Office-Supreme Court, U.S.
FILED
APR 15 1981
ALEXANDER L. STEVAS,
CLERK

80-1743

No. _____

IN THE

Supreme Court of the United States

OCTOBER TERM, 1980

CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK, DEPARTMENT OF PERSONNEL OF THE CITY OF NEW YORK, and THE NEW YORK CITY POLICE DEPARTMENT,

Petitioners,

—v.—

THE GUARDIANS ASSOCIATION OF THE NEW YORK CITY POLICE DEPARTMENT, INC., THE HISPANIC SOCIETY OF THE NEW YORK CITY POLICE DEPARTMENT, INC., NYDIA I. DIAZ, JAMES MICHAEL HIDALGO, WILFRED CEBELLERO, ANDRE LOPEZ, REINALDO SALGADO, DENISE SANTOS, DEBORAH HOLMES and PAMELA OBEY, individually and on behalf of all those similarly situated,

Respondents.

PETITION FOR A WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE SECOND CIRCUIT

ALLEN G. SCHWARTZ,
Corporation Counsel of
the City of New York,
Attorney for Petitioners,
100 Church Street,
New York, New York 10007.
(212) 566-4338/2183

LEONARD KOERNER,
JUDITH A. LEVITT,
of Counsel.

No. _____

IN THE

Supreme Court of the United States

OCTOBER TERM, 1980

CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK, DEPARTMENT OF PERSONNEL OF THE CITY OF NEW YORK, and THE NEW YORK CITY POLICE DEPARTMENT,

Petitioners,

—v.—

THE GUARDIANS ASSOCIATION OF THE NEW YORK CITY POLICE DEPARTMENT, INC., THE HISPANIC SOCIETY OF THE NEW YORK CITY POLICE DEPARTMENT, INC., NYDIA I. DIAZ, JAMES MICHAEL HIDALGO, WILFRED CEBELLERO, ANDRE LOPEZ, REINALDO SALGADO, DENISE SANTOS, DEBORAH HOLMES and PAMELA OBEY, individually and on behalf of all those similarly situated,

Respondents.

PETITION FOR A WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE SECOND CIRCUIT

QUESTIONS PRESENTED

1. Do bare statistics demonstrating a different pass rate for white and minority candidates on the written portion of Examination No. 8155, which tested for special abilities necessary to

perform the job of police officer, without any supporting evidence as to the relative qualifications of the white and minority candidates in the pool, give rise to a presumption of discrimination and satisfy plaintiffs' burden of establishing a *prime facie* case under Title VII of the Civil Rights Act of 1964, as amended?

2. Does the decision of the Court of Appeals, holding that Police Officer Examination No. 8155, although professionally developed and job-related, could not be used for rank-ordered selection of the best qualified candidates as mandated by the New York Constitution and New York State Civil Service Law, extend the scope of section 703(h) of Title VII beyond that intended by Congress, in conflict with this Court's decisions in *New York City Transit Authority v. Beazer,* 440 U.S. 568 (1979), and *United States* v. *South Carolina,* 445 F. Supp. 1094 (D.S.C. 1977), *summarily aff'd,* 434 U.S. 1026 (1978)?

3. Assuming that rank-ordered selection of candidates from Examination No. 8155 is not valid, did the Court of Appeals err in also holding that the pass score, set at the score of the 12,000th highest scoring candidate, determined by the projected manpower needs of the Police Department during the life of the eligible list, violates Title VII, because it may eliminate certain candidates who have qualifications equal to those of candidates who passed the examination?

TABLE OF CONTENTS

	PAGE
STATEMENT	1
OPINIONS BELOW	1
JURISDICTION	1
STATEMENT OF THE CASE	2
REASONS FOR GRANTING THE WRIT	6
1. In light of the ultimate burden upon plaintiffs to prove discrimination, plaintiffs' bare statistical analysis was insufficient to establish a *prima facie* case raising the presumption that Examination No. 8155 was discriminatory	8
2. The decision of the Court of Appeals invalidating the use of Examination No. 8155 for the selection of the best qualified candidates for the position of police officer as required by New York law far exceeds the scope of Title VII	10
CONCLUSION	14

TABLE OF AUTHORITIES

CASES:	PAGE
Albemarle Paper Co. v. *Moody*, 422 U.S. 405 (1975)	8, 11, 12
Castaneda v. *Partida*, 430 U.S. 428 (1977)	9
Commonwealth v. *O'Neill*, 465 F. Supp. 451 (E.D. Pa. 1979)	11
Firefighters Institute for Racial Equality v. *City of St. Louis*, 549 F.2d 506 (8th Cir.), *cert. den. sub nom. United States* v. *Banta*, 434 U.S. 819 (1977)	13
Furnco Construction Co. v. *Waters*, 438 U.S. 507 (1978)	8
Griggs v. *Duke Power Co.*, 401 U.S. 424 (1971)	6, 7, 10
Hazelwood School District v. *United States*, 433 U.S. 299 (1977)	9
International Brotherhood of Teamsters v. *United States*, 431 U.S. 324 (1977)	8, 9, 11
Jackson v. *Nassau County Civil Service Commission*, 424 F. Supp. 1162 (E.D.N.Y. 1976)	11
Jefferson v. *Hackney*, 406 U.S. 535 (1972)	8
Keyes v. *School Dist. No. 1*, 413 U.S. 189 (1973)	8
Kirkland v. *New York State Dept. of Correctional Services*, 520 F.2d 420 (2d Cir. 1975)	11
Louisville Black Police Officers v. *City of Louisville*, 20 F.E.P. Cases 1195 (W.D. Ky. 1979)	7
McDonnell Douglas Corp. v. *Green*, 411 U.S. 792 (1973)	6
National League of Cities v. *Usery*, 426 U.S. 833 (1976)	13

PAGE

New York City Transit Auth. v. *Beazer*, 440 U.S. 568 (1979) 8, 9, 10, 13

Scott v. *City of Anniston*, 597 F.2d 897 (5th Cir. 1979).. 13

Texas Dept. of Community Affairs v. *Burdine*, ____ U.S. ____, 49 U.S.L.W. 4214 (March 4, 1981) 8, 12

United States v. *City of Chicago*, 573 F.2d 416 (7th Cir. 1978) 7, 13

United States v. *Commonwealth of Virginia*, 454 F. Supp. 1077 (E.D. Va. 1978) 11

United States v. *South Carolina*, 445 F. Supp. 1094 (D.S.C. 1977), *summarily aff'd*, 434 U.S. 1026 (1978) 10, 13

United States v. *State of New York*, 21 E.P.D. ¶ 30, 314 (N.D.N.Y., Sept. 6, 1979) 7

Village of Arlington Heights v. *Metropolitan Housing Development Corp.*, 429 U.S. 252 (1977) 8

Vulcan Society v. *Civil Service Commission*, 490 F.2d 387 (2d Cir. 1973) 5

Washington v. *Davis*, 426 U.S. 229 (1976) 8, 10

STATUTES, RULES AND REGULATIONS:

Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. §§ 2000e, *et seq.* 1, 2, 4, 7, 8, 10, 12, 13

Section 703(h), 42 U.S.C. § 2000e-2(h) 6, 7, 11, 12

Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (1979)..... 4, 9

42 U.S.C. § 1983 ..... 1, 2

New York Civil Service Law 4, 10

§ 56 11

§ 61(1) 10

PAGE

CONSTITUTIONS:

United States Constitution

Tenth Amendment 13

Fourteenth Amendment 12

New York State Constitution 4, 10

Article V, § 6 10

OTHER AUTHORITIES:

Lerner, *Employment Discrimination: Adverse Impact, Validity, and Equality*, 1979 Supreme Court Review 17 7

Note, *Title VII and Public Employees: Did Congress Exceed Its Powers*, 78 COL. L. REV. 372 (1978) 13

STATEMENT

Petitioners, New York City Civil Service Commission, Department of Personnel and Police Department, seek a writ of certiorari to review a judgment of the United States Court of Appeals for the Second Circuit which affirmed in part and reversed in part an order of the United States District Court for the Southern District of New York and remanded for entry of a new decree.

Plaintiffs, black and Hispanic applicants for the position of New York City Police Officer, challenged the written portion of Police Officer Examination No. 8155 as discriminatory in violation of 42 U.S.C. § 1983 and Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. §§ 2000e, *et seq.* The Court of Appeals held that the test was job-related but that Title VII prohibited its use for rank-ordered selection of the best qualified candidates as mandated by the New York State Constitution and Civil Service Law.

OPINIONS BELOW

The opinion of the Court of Appeals, 630 F.2d 79 (2d Cir. 1980), is annexed as Appendix A. The order of the Court of Appeals, dated January 15, 1981, denying plaintiffs' petition for rehearing (not officially reported) is annexed as Appendix B. The revised opinion of the District Court, 484 F. Supp. 785 (S.D.N.Y. 1980), is annexed as Appendix C. The order of the District Court, entered January 11, 1980, is annexed as Appendix D.

JURISDICTION

The judgment of the Court of Appeals was entered on July 31, 1980. Plaintiffs' timely petition for rehearing was denied on January 15, 1981. This Court's jurisdiction is invoked under 28 U.S.C. §§ 1254(1) and 1331.

STATEMENT OF THE CASE

On June 30, 1979, defendants administered the written portion of Examination No. 8155 for the position of New York City Police Officer. Reflecting the success of an intensive recruitment drive aimed at attracting minority candidates, approximately one-third of the nearly 37,000 candidates who took the examination were black and Hispanic (App. 8a).* That examination resulted in the establishment of an eligible list based on test scores containing the names of approximately 13,000 candidates (App. 8a). Plaintiffs, black and Hispanic applicants, commenced this action on October 5, 1979, challenging the written test as discriminatory in violation of 42 U.S.C. § 1983 and Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. §§ 2000e, *et seq.*, and seeking declaratory and injunctive relief (A. 4-16).

After a trial on the merits, the United States District Court for the Southern District of New York rendered a decision** and order on January 11, 1980 (App. D), holding that the examination violated Title VII and was intentionally discriminatory. The District Court enjoined use of the list resulting from this examination unless 50% of the candidates appointed therefrom were black and Hispanic.

Defendants timely appealed to the Court of Appeals for the Second Circuit, and on July 31, 1980, that Court entered a judgment affirming in part and reversing in part the order of the District Court and remanding for entry of a new decree (App. A). The Court of Appeals reversed the finding of intentional discrimination but held that the use of the examina-

---

* Unless otherwise indicated, numbers in parentheses preceded by "App." refer to the Appendices annexed hereto, numbers preceded by "T." refer to the original trial transcript, and numbers preceded by "A." refer to the Appellants' Appendices submitted to the Court of Appeals.

** This decision was withdrawn and a revised opinion filed on January 23, 1980, during the pendency of defendants' appeal to the Court of Appeals, and after the appellants' brief had been filed in that Court (App. C).

tion for rank-ordered selection of the best qualified candidates violated Title VII.

As found by the Court of Appeals (App. 4a-7a), the job analysis and test preparation were thorough and done in a professional manner. Petitioners respectfully refer this Court to the opinion of the Court of Appeals for a detailed statement of the procedures followed by the New York City Department of Personnel and Police Department in developing Examination No. 8155.

Extreme caution was exercised throughout the test development procedures to ensure that the examination was fair and unbiased. Both in the job analysis and the test writing and review, all samples and panels of police officers were representative of the Police Department racially, ethnically and sexually (A. 33-84). The language in the examination questions was kept as simple as possible, and was, overall, on an eighth grade reading level so that vocabulary would not be an obstacle (App. 35a; T. 211; A. 359-62). In addition, extra time was given for completion of the test so that reading difficulties would not prevent candidates from competing equally (App. 35a).

Candidates were rank-ordered on the basis of their scores on the examination (App. 7a). The passing grade was set at the score of the 12,000th highest scoring candidate based upon the projected manpower needs of the Police Department over the course of the life of the eligible list (App. 8a).

Of the total candidates, 16.7% were identified as black, 14.2% were identified as Hispanic, 53.8% were identified as white, and 13.2% did not identify themselves. Of those who passed the examination, 7.6% identified themselves as black, 7.8% as Hispanic, and 66.6% as white (App. 8a).

Plaintiffs offered no statistics reflecting the ethnic composition of the 13.2% of the candidates who chose not to identify themselves (16.5% of those who passed the examination), nor evidence that the self-identification was accurate. No evidence was submitted concerning the relative qualifications, education

or motivation of the white and minority applicants who took the test. And no evidence was offered to establish that the qualifications necessary to become a police officer are uniformly present in the labor pool.

Based upon the analysis of the bare statistics under both the "four-fifths rule" established by the Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (1979), and standard deviation analysis, the Court of Appeals held that plaintiffs established a *prima facie* case of discrimination under Title VII. In so doing, the Court found that to require plaintiffs to substantiate the accuracy of the data submitted or to offer evidence of relative qualifications would be too onerous a burden upon plaintiffs (App. 13a, n.7).

The Court found, however, upon review of the job analysis and test preparation, that Examination No. 8155 was content valid and reasonably related to the job of police officer (App. 35a). Despite this finding of job-relatedness, the Court further held that there was not "a demonstration of such substantial test validity" (App. 37a) to support the use of the examination for rank-ordered selection for the position of police officer as required under the applicable provisions of the New York Civil Service Law, and, to the extent practicable, New York's Constitution (App. 35a-45a). The Court also held that the cut-off score—based upon the Police Department's projected manpower needs, rather than some concept of minimum competence—was invalid under Title VII (App. 45a-48a).

In reaching its conclusions regarding rank order and the pass score, the Court relied in significant part on its finding that one of the abilities necessary to perform the job of police officer, *i.e.*, human relations skills, identified as 30% of the job, was inappropriately tested for on a written examination. While it also indicated that if human relations skills had not been tested for in some manner, the test would have been unrepresentative of the job and therefore invalid, neither plaintiffs nor the Court offered any alternative method for testing for this ability (App. 30a-31a, 34a). Yet, based on this "inappropriateness," the Court found that defendants did not

prove that the examination was sufficiently job-related to demonstrate that one-point differences in test score would predict significant differences in job performance (App. 36a).* The Court also relied on the fact that significant numbers of candidates, including minority candidates, were bunched just below the cut-off score and within the range of the standard error of measurement (App. 46a-47a). Therefore, the Court concluded, the use of this cut-off score might exclude from employment certain candidates with qualifications equal to those who passed the examination (App. 46a-48a).

It was not disputed that plaintiffs' experts were at a loss to identify any other, more valid testing device which could have been used to select the most qualified candidates (T. 321-22, 389, 394, 496). Nor did plaintiffs offer any evidence that, had Examination No. 8155 been prepared in a more perfect manner, the disparate impact would have been reduced. Plaintiffs' expert, Dr. Richard Barrett, who, in his extensive experience in testing litigation, has never testified that an examination was valid, in response to questioning by the District Court, stated that he knew of no better test for the position of police officer than the one administered by defendants (T. 394):

> "THE COURT: Then let me take off. How would you construct a job-related test? Maybe I can get rid of the content.
>
> THE WITNESS: [Dr. Barrett]: I don't know how I would do it. The issue as posed in the guideline has two facets to

---

* The Court noted, in addition to the defects in the test preparation, the lack of evidence of reliability. In fact, the Court overlooked the evidence in the record that the test had a reliability coefficient of .93, admitted by plaintiffs' expert to be very high (T. 477; A. 68-69).

The other evidence of validity which the Court found lacking was either unavailable or unreliable, also as conceded by plaintiffs' expert (T. 367). A predictive validity study could not have been performed since only 415 of the highest ranking candidates had been appointed. A concurrent validity study would have been unreliable because of the experience and motivation of the incumbent police officers, and any pre-testing would have jeopardized the security of the examination. See *Vulcan Society* v. *Civil Service Commission*, 490 F.2d 387, 395 n.10 (2d Cir. 1973) (Friendly, J.).

> it, one is what test is used, and is there an alternative test to be used?
>
> And the other is, is there an alternative use of the same test? And I can't think of any alternative test. I have read literature and talked to people who have been developing them, and I have not been persuaded that these tests are any better than what we had here."

Despite this concession the Court of Appeals ignored the testimony of defendants' expert that evidence of validity of an examination indicates that higher scorers will, in general, perform better, and, therefore, not to use rank-ordered selection is to discard valuable information (T. 173-74, 225). Moreover, absent any evidence of feasible alternative selection devices or discriminatory intent, the Court nullified the mandate under New York law that, wherever practicable, candidates for public employment should be selected in rank order. It did this despite its acknowledgment that "[i]f a test is content valid [as this was held to be], it may be reasonable to infer that the test scores make some useful gross distinctions between candidates" and that "[c]andidates with high scores may well be expected to perform the job better than candidates with low scores" (App. 36a).

REASONS FOR GRANTING THE WRIT

Ten years ago, in the context of the use of an intelligence test to select employees "on the Company's judgment that [such tests] generally would improve the overall quality of the work force," this Court first interpreted that portion of section 703(h) of Title VII, 42 U.S.C. § 2000e-2(h), which authorizes the use of professionally developed tests to select employees. *Griggs* v. *Duke Power Co.*, 401 U.S. 424 (1971). In that case, this Court held that, upon the establishment of what *McDonnell Douglas Corp.* v. *Green*, 411 U.S. 792 (1973), would later characterize as a "*prima facie* case of discrimination," which there consisted of evidence of the disproportionate impact of

the test upon minority applicants and a history of intentional discrimination, Title VII requires an employer to demonstrate that an examination has a "manifest relationship to the employment in question"—that it is "job-related." *Griggs, supra,* at 413, 436.

In sharp contrast to *Griggs,* the Court of Appeals in this case has acknowledged that there is here no history of intentional discrimination nor any evidence of present discriminatory intent, and that the examination in question is a "second generation" selection procedure. That is, the City clearly sought to develop a test which meets professional standards and eliminates any elements of unfairness or bias.

In response to this "second generation" case, the Court of Appeals has radically extended *Griggs,* both with respect to plaintiffs' obligation to establish a *prima facie* case and defendants' burden of demonstrating job-relatedness. In so doing, the Court lost sight of the fact that the ultimate issue in any Title VII case is whether plaintiffs have proved discrimination. The Court effectively has read into section 703(h) the additional requirement that an employer utilizing a written employment examination to select best qualified applicants, which examination produces a "disparate impact," must demonstrate that its examination is the best and indeed more feasible than any examination currently in existence. This burden far exceeds that which petitioners believe was intended by Congress in enacting section 703(h) and in extending Title VII to the States and their subdivisions in 1972, or which this Court ever contemplated in *Griggs.** The far-reaching effects of this decision on the entire public employment sector, which is, by and large, under the mandate of civil service laws, and on the

---

* The Second Circuit is not alone in requiring employers to demonstrate that they have used the most perfect test. See, *e.g., United States* v. *City of Chicago,* 573 F.2d 416 (7th Cir. 1978); *Louisville Black Police Officers* v. *City of Louisville,* 20 FEP Cases 1195 (W.D. Ky. 1979); *United States* v. *State of New York,* 21 E.P.D. ¶30,314 (N.D.N.Y., Sept. 6, 1979). See, also, Lerner, *Employment Discrimination: Adverse Impact, Validity, and Equality,* 1979 Supreme Court Review 17.

private sector, require review and clarification by this Court of the scope and obligations of Title VII in a "disparate impact" case.

1. In light of the ultimate burden upon plaintiffs to prove discrimination, plaintiffs' bare statistical analysis was insufficient to establish a *prima facie* case raising the presumption that Examination No. 8155 was discriminatory.

This Court has clearly stated that the ultimate burden of proving discrimination is on the plaintiffs. See *New York City Transit Auth.* v. *Beazer,* 440 U.S. 568, 587, n.31 (1979). See, also, *Albemarle Paper Co.* v. *Moody,* 422 U.S. 405, 425 (1975). The "division of intermediate evidentiary burdens bring the litigants and the court expeditiously and fairly to this ultimate question." *Texas Dept. of Community Affairs* v. *Burdine,* ____ U.S. ____, 49 U.S.L.W. 4214, 4216 (March 4, 1981). The first of these intermediate burdens is the establishment of a *prima facie* case which creates a presumption of discrimination. See *McDonnell Douglas Corp.* v. *Green, supra.* The evidence necessary to satisfy plaintiff's burden must be sufficient to eliminate the most common nondiscriminatory reasons for the disparate results of the selection device used by the employer. See *International Brotherhood of Teamsters* v. *United States,* 431 U.S. 324, 358 and n.44 (1977); *Furnco Construction Co.* v. *Waters,* 438 U.S. 507, 577 (1978).

In a wide variety of contexts this Court has cautioned against reliance on bare statistics alone to establish a *prima facie* case, finding that in many instances, to recognize the limited probative value of disproportionate impact is merely to acknowledge 'the heterogeneity of the nation's population.' " *Village of Arlington Heights* v. *Metropolitan Housing Development Corp.,* 429 U.S. 252, 266 n.15, quoting *Jefferson* v. *Hackney,* 406 U.S. 535 (1972). See *Washington* v. *Davis,* 426 U.S. 229 (1976) (employment examination); *Keyes* v. *School Dist. No. 1,* 413 U.S. 189 (1973) (school segregation). Yet it was precisely this type of bare analysis that was accepted by the Court of Appeals in this case.

Here, plaintiffs' sole evidence to establish their *prima facie* case was applicant flow data indicating that of the candidates who chose to identify themselves racially or ethnically, white candidates passed the written portion of Examination No. 8155 at a higher rate than minority candidates. The total 'ack of evidence of qualifications renders meaningless both plaintiffs' comparison of white and minority pass rates in terms of the "four-fifths rule," as well as the application of the standard deviation analysis approved by this Court in *Castaneda* v. *Partida,* 430 U.S. 428 (1977). In *Hazelwood School District* v. *United States,* 433 U.S. 299, 308 n.13 (1978), this Court specifically cautioned that "[w]hen special qualifications are required to fill particular jobs, comparison to the general population (rather than to the smaller group of individuals who possess the necessary qualifications) may have little probative value." This is particularly true where, as here, the employer has conducted an intensive recruitment effort directed at attracting protected group members. Uniform Guidelines on Employee Selection Procedures, sec. 4D. While it may not be possible for plaintiffs to refute every circumstance that could explain the disproportionate impact shown by the bare statistics, it is required, contrary to the ruling of the Court of Appeals, that plaintiffs eliminate the most common nondiscriminatory factors.

Furthermore, while applicant flow data is relevant in evaluating allegations of employment discrimination, standing alone such data is insufficient to establish a *prima facie* case. In *Hazelwood, supra,* this Court held that such data must be evaluated in light of all other statistical and nonstatistical evidence to determine their usefulness. Accord, *Teamsters, supra; Beazer, supra.*

Particularly in the absence of a history of intentional discrimination, plaintiffs cannot rely on the rigid application of statistical analyses to overcome their burden of establishing a *prima facie* case of discrimination. There is no such history in the City of New York and this test did not result in the exclusion of minority candidates. Under these circumstances, and in accord with the approach adopted by this Court in

Beazer, supra, the evidence utterly fails to support any presumption of discrimination.

2. The decision of the Court of Appeals invalidating the use of Examination No. 8155 for the selection of the best qualified candidates for the position of police officer as required by New York law far exceeds the scope of Title VII.

Under any fair reading of Title VII it must be held that defendants satisfied their burden of demonstrating that Examination No. 8155 has a "manifest relationship" to the job of police officer. *Griggs, supra.* Yet, the Court of Appeals, although finding the examination to be job-related, invalidated rank-ordered selection of the best qualified candidates and the cut-off score based upon projected manpower needs on the theory that the examination could have been prepared in a more perfect fashion. This Court has never construed Title VII in this manner, imposing upon an employer the obligation to prove that the selection device used is the best that could have been devised. See *New York City Transit Auth.* v. *Beazer, supra; United States* v. *South Carolina,* 445 F. Supp. 1094 (D.S.C. 1977), *summarily aff'd,* 434 U.S. 1026 (1978). *Cf. Washington* v. *Davis,* 426 U.S. 229 (1976). If permitted to stand, the decision of the Court of Appeals will effectively preclude the City from seeking to hire the most qualified employees for public service as required by the New York State Constitution and Civil Service Law.

It cannot be disputed that the City's interest in employing the candidates best qualified for positions within the public service is legitimate and nondiscriminatory. This public policy is embodied in both the New York State Constitution and Civil Service Law § 61(1). Article 5, § 6 of the New York Constitution provides in pertinent part:

> "Appointments and promotions in the Civil Service of the state and all of the civil divisions thereof, including cities and villages, shall be made according to merit and fitness

> to be ascertained, as far as practicable, by examination which, as far as practicable shall be competitive. . . ."

Rank-ordered selection ensures that the most qualified candidates will enter public service. Rank-ordered selection is particularly important in this regard when considering that an eligible list may have a life of four years under Civil Service Law § 56. It is likely that, if rank order is discarded, after a year or more from the giving of the examination, candidates who are better qualified and have not been reached for appointment will have found other positions and lost interest in public employment. Rank-ordered selection also serves the important interest of objectivity in public employment decisions. And, in combination with a cut-off score based on projected manpower needs, it reduces the public safety and cost risks inherent in hiring unqualified or minimally qualified employees.

The courts, including the Court of Appeals for the Second Circuit, have traditionally given great deference to these goals and the aims of the civil service law. See *Kirkland* v. *New York State Dept. of Correctional Services,* 520 F.2d 420, 428 (2d Cir. 1975), *cert. den.,* 429 U.S. 823 (1976); *Jackson* v. *Nassau County Civil Service Commission,* 424 F. Supp. 1162, 1166 (E.D.N.Y. 1976); *Commonwealth* v. *O'Neill,* 465 F. Supp. 451, 463-65 (E.D. Pa. 1979); *United States* v. *Commonwealth of Virginia,* 454 F. Supp. 1077, 1101 (E.D. Va. 1978). Indeed, the concept of merit selection is embodied in the language of section 703(h), which protects from challenge a "*bona fide* seniority or merit system" unless designed to discriminate. See *International Brotherhood of Teamsters* v. *United States, supra.*

The burden imposed by the Court of Appeals, to prove that one-point differentials in test scores will accurately predict demonstrable differences in job performance, is, simply, insurmountable. It must be kept in mind that no pre-employment test approaches such accuracy of prediction. *Albemarle Paper Co.* v. *Moody,* 422 U.S. 405, 449 (1975) (Blackmun, J.,

concurring). This is not surprising since the psychological standards for employee selection are constantly evolving as evidenced by the series of guidelines issued by the EEOC which are frequently contradictory. See *Albemarle Paper Co., supra,* at 451-52 (Burger, C.J., concurring in part and dissenting in part). Neither the Court nor the plaintiffs in this case were able to point to any alternative selection device which would meet the legitimate employment goals of the City and which would have resulted in a lesser disparate impact. Absent such evidence, the City cannot be required to abandon these goals in favor of seeking mere minimum competency in its employees in order to achieve proportionate minority representation.

Most recently, in the context of a "disparate treatment" case, this Court stated its view of the substantive limitations of Title VII:

> "The statute was not intended to 'diminish traditional management prerogatives.' [*Steelworkers* v. *Weber,* 433 U.S. 193, 207 (1979)]. It does not require the employer to restructure his employment practices to maximize the number of minorities and women hired. *Furnco Construction Co.* v. *Waters,* 438 U.S. [567, 577-78 (1978)]. . . .
>
> Rather, the employer has the discretion to choose among equally qualified candidates, provided the decision is not based upon unlawful criteria. The fact that a court may think that an employer misjudged the qualifications of the applicants does not in itself expose him to Title VII liability. . . ."

Texas Dept. of Community Affairs v. *Burdine, supra,* ____ U.S. ____, 49 U.S.L.W., at 4217. Nothing in section 703(h) of Title VII suggests that Congress, in authorizing the use of professionally developed tests, intended any greater restrictions upon the employer.

Certainly, the Congress which extended Title VII to the States and their subdivisions in 1972 under the authority of the Fourteenth Amendment, mindful of the principles of federal-

ism found in the Tenth Amendment, could not have intended the restriction upon the City's ability to govern itself in an effective manner which results from the decision of the Court of Appeals in this case. See, generally, *National League of Cities* v. *Usery,* 426 U.S. 833 (1976); *Scott* v. *City of Anniston,* 597 F.2d 897 (5th Cir. 1979); *United States* v. *City of Chicago,* 573 F.2d 416 (7th Cir. 1978); *Firefighters Institute for Racial Equality* v. *City of St. Louis,* 549 F.2d 506 (8th Cir.), *cert. den. sub nom. United States* v. *Banta,* 434 U.S. 819 (1977); Note, *Title VII and Public Employees: Did Congress Exceed Its Powers,* 78 Col. L. Rev. 372 (1978). Although this constitutional tension was not an articulated ground for this Court's holdings in *Beazer, supra,* and *United States* v. *South Carolina, supra,* we submit that it affords some explanation for the analytic approach there adopted.

The decision of the Court of Appeals in this case places an enormous, and, indeed, impossible burden on both public and private employers, unrelated to the goal of Congress in enacting Title VII to ensure minority group members equal access to employment opportunities. Petitioners urge this Court to clarify the obligations imposed upon employers by Title VII in the context of this "disparate impact" case.

CONCLUSION

For these reasons, a writ of certiorari should be granted.

April 15, 1981.

Respectfully submitted,

ALLEN G. SCHWARTZ,
Corporation Counsel of
the City of New York,
Attorney for Petitioners.

LEONARD KOERNER,
JUDITH A. LEVITT,
of Counsel.

Appendix A

Opinion of the Second Circuit

UNITED STATES COURT OF APPEALS

FOR THE SECOND CIRCUIT

No. 849—September Term, 1979

(Argued February 6, 1980 Decided July 31, 1980)

Docket No. 80-7027

THE GUARDIANS ASSOCIATION OF THE NEW YORK CITY POLICE DEPARTMENT, INC., THE HISPANIC SOCIETY OF THE NEW YORK CITY POLICE DEPARTMENT, INC., NYDIA I. DIAZ, JAMES MICHAEL HIDALGO, WILFRED CEBELLERO, ANDRE LOPEZ, REINALDO SALGADO, DENISE SANTOS, DEBORAH HOLMES and PAMELA OBEY, individually and on behalf of all those similarly situated,

Plaintiffs-Appellees,

—v.—

CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK, DEPARTMENT OF PERSONNEL OF THE CITY OF NEW YORK, and THE NEW YORK CITY POLICE DEPARTMENT,

Defendants-Appellants.

Before:

MANSFIELD and NEWMAN, *Circuit Judges,*
and SIFTON,* *District Judge.*

* Of the United States District Court for the Eastern District of New York, sitting by designation.

Appeal from a decision by the United States District Court for the Southern District of New York (Robert L. Carter, Judge), 484 F. Supp. 785, holding that New York City violated Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e-2, in administering an examination to applicants for positions in the New York City Police Department.

Affirmed in part, vacated in part, and remanded.

PETER BIENSTOCK, New York, N.Y. (M.D. Taracido, Kenneth Kimmerling, Robert L. Becker, Puerto Rican Legal Defense & Education Fund, New York, N.Y., on the brief), *for plaintiffs-appellees.*

L. KEVIN SHERIDAN, New York, N.Y. (Allen G. Schwartz, Corporation Counsel, Judith A. Levitt, Steven M. Goldberg, Maureen M. McCabe, New York, N.Y., on the brief), *for defendants-appellants.*

DAVID L. ROSE, Washington, D.C. (Robert B. Fiske, Jr., U.S. Atty., Nancy E. Friedman, Richard N. Papper, Dennison Young, Jr., Asst. U.S. Attys., New York, N.Y., Drew S. Days III, Asst. Atty. Gen., Steven H. Rosenbaum, Washington, D.C., on the brief), *for the United States as amicus curiae.*

(Ira H. Leibowitz, Barry Lasky, Garden City, N.Y., submitted a brief for The Policewomen's Endowment Assoc., Inc., as *amicus curiae.*)

(H. Elliot Wales, New York, N.Y., submitted a brief for Seven Civil Service Organizations as *amicus curiae.*)

NEWMAN, *Circuit Judge:*

This employment discrimination suit pursuant to Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-2, once again requires this Court to venture into the complex realm of testing and test validation. The test at issue was designed by New York City officials and administered on June 30, 1979 to 36,797 applicants for positions on the City's police force. Plaintiffs are the Guardians Association of the New York City Police Department, Inc., an organization of Black police officers, the Hispanic Society of the New York City Police Department, Inc., an organization of Hispanic police officers, and eight individual Black or Hispanic applicants. Defendants are the New York City Department of Personnel, which performed much of the test preparation, the New York City Civil Service Commission, and the New York City Police Department. The United States District Court for the Southern District of New York (Robert L. Carter, Judge) found that use of the test unjustifiably discriminates against Blacks and Hispanics in violation of Title VII. *Guardians Ass'n v. Civil Service Commission,* 484 F. Supp. 785 (S.D.N.Y. 1980). The Court ordered a broad remedy, including a 50% minority hiring quota. We affirm the District Court's finding that the City's specific use of the test violates Title VII, but vacate the remedy and remand for entry of a revised decree.

I. Factual Background

The test in question, designated Exam No. 8155, was designed to select candidates for hiring as entry-level police officers. Those who pass the exam are selected, in rank order of their test scores, to complete the other aspects of the hiring process—a medical examination, a physical agility test, a psychological test, and a character investigation. These last four components of the hiring process are scored only on a pass/fail basis. Thus, an appellant's score on Exam No. 8155 is a major determinant of his prospects for becoming a police officer. It is also the only feature of the process alleged to have

a discriminatory impact. Once an applicant scores high enough to be selected for the final four hiring steps and successfully completes those steps, he or she becomes a sworn police officer and enters the police academy for five months of training. While successful completion of the training program is a requirement of continuing as a police officer, the Department does not use the training program as a selection device, but anticipates that nearly all academy entrants will go on to active duty.

The exam was developed by a fairly elaborate two-stage process; the first stage was an analysis of the police officer's job, and the second was construction of the test itself. The job analysis consisted of five separate steps. First, the Department of Personnel identified 71 tasks that police officers generally perform, based on interviews with 49 police officers and 49 supervisors. Second, a panel of seven officers and supervisors reviewed the list to add any tasks that had been omitted, and to eliminate those items that were duplicative, or too specialized to be performed by entry-level officers. The result was a consolidated list of 42 entry-level tasks.

Third, a questionnaire was distributed to 5,600 police officers, requesting them to rate each of the 42 tasks on the basis of its frequency of occurrence, its importance, and the amount of time normally spent in performing it. The 2,600 responses that were received were then analyzed by computer to yield a ranking of the 42 tasks, according to the combined rating of all the responses. In addition, faculty members of John Jay College were asked to observe police officers during an entire tour of duty and record the tasks that they performed; their survey generally confirmed the identification of the 42 tasks.

In the fourth step of the job analysis, the Department of Personnel divided the list of 42 ranked tasks into clusters of related activities. Five such clusters were established: the arrest process, providing assistance to people, police operations, stationhouse activities, and handling unusual and other occurrences. The fifth step was an analysis of all five clusters, each one by a separate panel of police officers, to identify the "knowledge, skills and abilities" required to perform these

tasks at the entry level, and to assign percentages reflecting the relative importance of each of the identified knowledges, skills, and abilities for the cluster as a whole. One panel listed five such qualities for its cluster, all of which are properly characterized as "abilities" or "skills" (hereafter referred to as "abilities"): recalling facts, filling out forms, understanding and applying statutory definitions of crimes, understanding written instructions and applying appropriate procedures, and human relations skills, including communication techniques. Each of the other four panels used the first panel's list of abilities, but developed its own percentages to express the relative importance of each ability to the tasks within its cluster.

The second major stage in developing Exam No. 8155, the process of test construction, consisted of four identifiable steps. First, the percentages of the five abilities necessary to perform each of the five task clusters were multiplied by the weightings that had been given to each task in Step 3 of the job analysis on the basis of frequency, importance, and time spent. This yielded a general measurement for the importance of each of the five abilities for performance of the job of police officer. As a result of this computation, the Department of Personnel concluded that on a test with 100 questions, 15 questions should test for the ability to recall facts, 9 questions for filling out forms, 14 questions for understanding and applying sections of the criminal law, 32 questions for understanding written instructions and applying appropriate procedures, and 30 questions for human relations skills. Next, a group of eleven police officers was selected to write multiple-choice questions that tested for the five abilities, as they related to the 42 identified tasks. The officers wrote many of these questions from Police Academy materials and similar sources, however, without having access to descriptions of the five identified abilities, or the 42 ranked tasks. In the third step, Department of Personnel staff members who did have access to the description of abilities and the ranking of tasks reviewed the questions written by the police officers to assure that the questions were not ambiguous, overly complex, overly specialized, or dependent on prior knowledge. As a result of this review, some

questions were discarded, others were revised, and still others were added. Finally, the resulting questions were subjected to a further review by a panel of six police experts, and by various members of the Department of Personnel.

The test that resulted consisted of 100 multiple-choice questions, designed so that the candidate could answer correctly without knowledge of any information beyond what was provided on the test itself. The test materials were determined by the Department of Personnel to require an eighth-grade reading level, on the average, although the 14 questions on law required college-level reading ability. The estimated time for completing the exam was 1½ hours, but 3½ hours were allowed.

The first part of the exam, designed to measure the ability to recall facts, consisted of a page-and-a-half description of a burglary, and a series of 15 questions to be answered without referring back to the description. In the second part, testing ability to fill out forms, the candidates were given a simplified arrest form, and a page-long description of both a robbery and an arrested suspect, and then asked 9 questions about the proper entries to be made in filling out the form. Part three, intended to test ability to apply provisions of law, consisted of 14 questions, each briefly presenting the facts of an incident, and then requiring the candidate to identify the precise criminal offense involved on the basis of definitions provided in the test materials.[1] The remaining 62 questions, of which 32

---

1 One example:

37. Chris Hart and Larry Burns are walking by a big museum late one night. Hart notices that, although the museum is closed to the public, one of the doors is unlocked. Hart suggests that for a prank they go into the museum to see the exhibits. Hart has a flashlight with him while Burns has an illegal gun hidden under his clothing. Hart does not know that Burns has a gun illegally in his possession. About five minutes after they enter the museum, they hear footsteps and leave the museum the same way they entered. According to the definitions given,

(A) Burns committed the crime of criminal trespass, but Hart did not

(B) Hart committed the crime of criminal trespass, but Burns did not

were intended to measure the ability to follow appropriate procedures and 30 were intended to measure human relations skills, consisted of general instructions as to procedures or appropriate responses for certain types of situations, a description of a specific situation, and then one or more questions asking the proper response to the situation presented. Three of the questions dealing with appropriate procedures, for example, involved the proper response to a bomb threat. Four of the questions in the human relations section involved the proper way to deal with a person who appears to be mentally ill.

The test was scored from zero to one hundred, with one point given for each correct answer, and bonus points given for veterans.[2] The candidates were then rank-ordered on the basis of their scores. Scores were generally high, with 13% of the applicants scoring 98 or above, and fully 50% scoring 91 or above. Because of the number of candidates taking the test and the bunching of candidates in the upper range of scores, each point a candidate achieved made a substantial difference in his

---

(C) both Hart and Burns committed the crime of criminal trespass

(D) neither Hart nor Burns committed the crime of criminal trespass.

The definition provided for criminal trespass is as follows:

> The crime of criminal trespass is committed when a person knowingly enters or remains in a building in which he has no right to be and while in the building, possesses, or knows that another person accompanying him possesses, an explosive or a gun.

It will be noted, simply as an indication of the difficulty of the test construction enterprise, that there is a slight ambiguity in this question. The applicant is told that Hart is unaware that "Burns has a gun illegally in his possession," but not whether Hart was unaware that Burns had a gun at all. This could affect the answer, since, according to the definition given, knowledge of any gun, even of a legal one, would render Hart's action a criminal trespass. While it would be unreasonable to suggest that a test must be free of every possible ambiguity in order to be acceptable, the ease with which such ambiguities can appear emphasizes the value of confirming the test's reliability by some empirical procedure. *See* Section IV *infra*.

2 As a result of these bonus points, it was possible to achieve a score as high as 110. Some 482 applicants, or 1.3% of those taking the test, scored above 100.

position on the rank-ordering list. More than 2,000 applicants achieved a score at each numerical grade from 92 to 97.

The passing grade was determined in the following manner. The Police Department first estimated that 4,000 police officers would be hired during the four-year period for which the eligibility list resulting from Exam No. 8155 would be valid. The Department further estimated that only one out of three applicants who passed Exam No. 8155 would successfully complete all the remaining steps in the hiring process. Therefore, if this eligibility list was to meet the Department's needs, 12,000 persons had to pass the exam to provide the 4,000 needed police officers. With all this in mind, the Department simply set the passing grade at the score achieved by the 12,000th highest scoring candidate, which turned out to be 94. Because of the bunching phenomenon, a large number of candidates, 2,124, achieved this same score, so that the actual number who received a passing grade was 13,749.

Of the 36,797 applicants who took the test, 6,142 identified themselves as Black, 5,239 identified themselves as Hispanic, 19,798 identified themselves as white, and 4,847, or 13.2% did not identify their race. Thus, identified Blacks constituted 16.7% of the total applicants, and 19.7% of all those who identified their race, while the equivalent percentages for identified Hispanics were 14.2% and 16.8%, and for identified Whites, 53.8% and 64.5%.

Of those who passed the exam, *i.e.*, scored 94 or better, 7.6% had identified themselves as Black and 7.8% had identifed themselves as Hispanic, for a known minority population in the passing group of 15.4%, against a known minority population in the applicant pool of 30.9%. In contrast, 66.6% of the passing applicants had identified themselves as White, although Whites comprised only 53.8% of the applicant pool.

Viewed in another and more revealing way, the figures show that, among those who had identifed themselves by race, the passing rate for Whites was 45.9% compared to 17% for Blacks and 20.5% for Hispanics. The combined minority pass rate was thus about two-fifths of the pass rate for Whites.

The Police Department accepted 415 candidates from the list in November, 1979 and planned to hire another 380 in January, 1980. Of this group of 795 candidates, 89.2% were White, 3.5% were Black, and 6.8% were Hispanic. The selection rates (number chosen compared to number of applicants) for these first two uses of the list were 0.5% for Blacks, 1% for Hispanics, and 3.6% for Whites.

II. The District Court's Prior Proceedings and Decision

The plaintiffs filed their complaint in this suit, together with a motion for a preliminary injunction, in October, 1979, before any candidates had been accepted on the basis of the list. They charged that the intended use of the lists by the Police Department constituted discrimination against Blacks and Hispanics in violation of the Fourteenth Amendment, Title VII, and various other Federal and state laws. The District Court, by consent of the parties, consolidated the hearing on the preliminary injunction and the trial on the merits. This proceeding was held on November 13, 14, and 15, 1979, shortly after the Police Department's first use of the list. On January 11, 1980, three days before the Department intended to use the list to accept a second group of trainees, the District Court held a second hearing. That same day, the Court issued an opinion, which was subsequently re-issued in revised form on January 23.[3]

---

3 By the time of the November hearing, the Police Department had already proceeded to use the list to accept 415 trainees, as described above. In order to forestall any further hiring on the basis of the list, and to assist the City in making alternative arrangements, the District Court informed the parties on December 17, 1979 that the test violated Title VII, and orally enjoined the defendants from its further use. On December 27, 1979, the City filed an order to show cause and a motion for a stay, in which it stated that the Police Department needed to accept an additional 380 trainees from the list on January 14, 1980. When the Court denied its motion, the City filed a petition in this Court for a writ of mandamus. This Court granted the petition, ordering the District Court to issue findings of fact and conclusions of law, pursuant to Fed. R. Civ. P. 52(a), at least 48 hours before any injunction against the City was to take effect. The District Court then

The Court's basic conclusion was that Exam No. 8155 violated Title VII. In reaching this conclusion, the Court used the common mode of Title VII analysis, in which the plaintiff is first required to establish a *prima facie* case on the basis of disparate impact, and then the defendant is required to rebut the plaintiff's case by proving that the disparity results from legitimate, job-related selection procedures. The Court first found that the disparity between the percentage of minority group members who achieved a passing score and the percentage of minority groups members in the applicant pool was sufficient to establish a *prima facie* case. It based this finding of disparate impact on the standards developed by the Supreme Court in *Castaneda v. Partida,* 430 U.S. 482 (1977), and by the Equal Employment Opportunity Commission (EEOC) in its Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (1979). *Castaneda* stated that, in cases involving large samples, "if the difference between the expected value [from a random selection] and the observed number is greater than two or three standard deviations," a *prima facie* case is established. 430 U.S. at 496 n.17.[4] The Uniform Guidelines provide that "[a] selection rate for any

---

held its second hearing, which was devoted to the issue of relief. Since the Court's decision, issued that day, preceded the January 14 action by more than 48 hours, it fulfills the requirements of this Court's mandamus.

4 The standard deviation for a particular set of data provides a measure of how much the particular results of that data differ from the expected results. In essence, the standard deviation is a measure of the average variance of the sample, that is, the amount by which each item differs from the mean. The number of standard deviations by which the actual results differ from the expected results can be compared to the normal distribution curve, yielding the likelihood that this difference would have been the result of chance. The likelihood that the actual results will fall more than one standard deviation beyond the expected results is about 32%. For more than two standard deviations, it is about 4.6% and for more than three standard deviations, it is about .03%. On this basis, the Supreme Court concluded in *Casteneda* that when actual results fell more than three standard deviations from the expected result (that is, a race-neutral selection), the deviation could be regarded as caused by some factor other than chance.

race, sex, or ethnic group which is less than four-fifths ($^4/_5$) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact." § 1607.4(D) [hereinafter Guideline sections are cited only by the subdivisions of 29 C.F.R. § 1607]. The District Court then noted that the discrepancy between the percentage of minority group members in the applicant pool and the percentage of minority groups members who passed the test was 39 standard deviations. The evidence also showed that the passing rate of the minority group members was 44.3% of the passing rate of Whites, or about two-fifths.

Having concluded that the plaintiffs had established a *prima facie* case, the Court next concluded that the test was not sufficiently valid to constitute a legitimate attempt to choose those applicants who would become better police officers. Relying primarily on the EEOC Guidelines, the Court stated that a determination of validity based on the content of the test would be inappropriate, first because the test purported to measure abilities that the accepted applicants would be trained to acquire, see Guidelines, § 14(C)(1), and second, because the test actually measured constructs, not abilities, see *id.*[5] Moreover, the Court concluded that the job analysis was not sufficiently precise to satisfy the Guideline requirement even for content validation, see Guidelines § 14(C)(2). Since content validation was the only method of validation which the City attempted, the Court concluded that the test was invalid, and thus an inadequate rebuttal to the plaintiffs' *prima facie* case.

In fashioning relief, the District Court noted that a previous examination administered by the New York City Police Depart-

---

5 A construct is generally defined as "an idea developed or 'constructed' as a work of informed, scientific imagination; that is, it is a theoretical idea developed to explain and to organize some aspects of existing knowledge." American Psychological Association, Inc., Standards for Educational & Psychological Tests 29 (1974) [hereinafter APA Standards]. Neither the APA Standards nor the Guidelines appear to include a definition of content, apart from the concept of content validity, although the Guidelines do describe content as involving "knowledges, skills, or abilities." Guidelines § 14(C)(1).

ment had been found to be in violation of Title VII. *Guardians Association v. Civil Service Commission,* 431 F. Supp. 526 (S.D.N.Y.), *vacated and remanded on other grounds,* 562 F.2d 38 (2d Cir. 1977). Concluding that the defendants had "persisted in devising and utilizing testing procedures that continue to discriminate against blacks and hispanics," the District Court found that the defendants' "studied adherence to discriminatory procedures at this point must be deemed conscious and deliberate." 484 F. Supp. at 798. On this basis, the Court held that "[a]ffirmative action is mandated as an interim measure either until such discrimination has been totally eliminated or until defendants proceed to select police officers under procedures that are in full compliance with Title VII." *Id.* at 27. It enjoined the City from using Exam No. 8155, although permitting it to use the eligibility list from that exam for purposes designated by the Court. In its order, also issued on January 11, the Court ordered the Police Department to achieve at least 30% minority composition of the force, a level comparable to the percentage of minorities in the labor force of the relevant hiring area. To achieve this goal, the Court further ordered that the defendants should "as an interim goal appoint 50% of their entry level police officers from among qualified black and hispanic applicants." Finally, the Court awarded the plaintiffs attorneys' fees and costs, and retained jurisdiction "for such further relief or other orders as may be necessary or appropriate to enforce and insure rights to equal employment opportunity within the New York City Police Department."

The City moved to stay the District Court's order pending consideration of its appeal from the decision. This Court denied the motion. However, it granted a conditional stay, in view of the City's declared need to hire new policy officers[6] and set an expedited schedule for the appeal.

---

6 The terms of this conditional stay were that, if the City wished to hire, pending appeal, it should establish two pools of candidates, one consisting of all the minority applicants who passed the exam, and the second consisting of all others who passed, and select trainees from these pools in the ratio of one minority applicant for every two others. The 415 applicants already hired were to be counted in determining whether the new hires conformed to this 1 to 2 ratio.

III. The Framework of Title VII Analysis

As the District Court concluded, the accepted procedure for Title VII cases is to require the plaintiffs to establish a *prima facie* case, and then to require the defendants to rebut this showing with proof that the test was legitimately job-related. *See Albemarle Paper Co. v. Moody,* 422 U.S. 405 (1975); *McDonnell Douglas Corp. v. Green,* 411 U.S. 792 (1973); *Griggs v. Duke Power Co.,* 401 U.S. 424 (1971). The Court correctly concluded that a *prima facie* case had been established. By any reasonable measure, including the standard deviation rule of *Castaneda, supra,* or the four-fifths rule of the EEOC Guidelines, Exam No. 8155 had a disparate racial impact.

The City argues that statistics alone, specifically a comparison of the racial composition of the passing group to that of the applicant group, are not sufficient to establish a *prima facie* case. But statistics showing a significantly disparate racial impact have consistently been held to create a presumption of Title VII discrimination. See *International Brotherhood of Teamsters v. United States,* 431 U.S. 324, 339 (1977); *United States v. City of Chicago,* 549 F.2d 415, 428 (7th Cir.), *cert. denied,* 434 U.S. 875 (1977); *Kirkland v. New York State Department of Corrections,* 520 F.2d 420, 425 (2d Cir. 1975), *cert. denied,* 429 U.S. 823 (1976); *Vulcan Society v. Civil Service Commission,* 490 F.2d 387, 392-93 (2d Cir. 1973); *Chance v. Board of Examiners,* 458 F.2d 1167, 1172-73 (2d Cir. 1972).[7]

---

7 The City hypothesizes some situations in which statistics could be misleading (*e.g.*, if some of the candidates taking the test had not been eligible to apply for the position) but presents no evidence to show that this occurred. To accept such unsupported possibilities, and require the plaintiffs to refute every circumstance that could explain the disparate impact shown by the statistics, would create an onerous burden of proof, far in excess of the Title VII standards as interpreted by the Supreme Court. *See Dothard v. Rawlinson,* 433 U.S. 321, 329-30 (1977); *Jones v. New York City Human Resources Administration,* 528 F.2d 696, 698 (2d Cir. 1976).

The City also claims that finding a *prime facie* Title VII violation by state or local governments without a showing of discriminatory intent violates the Tenth Amendment. This view has been definitively rejected by the Seventh Circuit in *United States v. City of Chicago,* 573 F.2d 416, 422-24 (7th Cir. 1978), and we agree with that analysis. Congress may enforce the Fourteenth Amendment by legislation that prohibits practices the Amendment might not of its own force condemn. See *Katzenbach v. Morgan,* 384 U.S. 641 (1966).

The real issue in this case, therefore, is whether the defendants have rebutted the plaintiffs' *prima facie* case by proving that its test was job-related: that the test accurately selected applicants who would be better police officers. Adjudication of this issue presents a more complex problem in the present case than it has in many previous Title VII suits. Many of the previous suits involved tests that were so artlessly constructed that they could be judged invalid without extensive inquiry, fine distinctions, or a precise notion of where the line between validity and invalidity was located. See, *e.g., Griggs, supra,* 401 U.S. at 431 (intelligence tests used "on the Company's judgment that they generally would improve the overall quality of the work force"); *United States v. N.L. Industries, Inc.,* 479 F.2d 354, 371 (8th Cir. 1973) (test given to one applicant "consisted of four or five mathematical problems which a Company employee jotted down on a sheet of yellow paper"); *Brito v. Zia Co.,* 478 F.2d 1200, 1205-06 (10th Cir. 1973) (test based almost entirely on subjective judgments of supervisors, not administered or scored under controlled and standardized conditions); *Vulcan Society, supra,* 490 F.2d at 396-98 (no job analysis, test measured abilities that were clearly of secondary importance to job).

Exam No. 8155, in contrast, is a "second generation" selection procedure. Despite the various flaws in construction of the test, it is clear that some attempt was made to develop the test with recognition of at least some of the standards that courts have established in the first wave of Title VII cases. Aware that the validity of the test would likely have to be demonstrated, the City performed an extensive job analysis,

consciously used Guideline concepts in determining the qualities that were being tested for, and attempted to eliminate extraneous variables, such as the applicant's prior knowledge, his reading level, and his ability to complete the test in a relatively short amount of time.

Nevertheless the plaintiffs have alleged and the District Court has concluded that the construction and use of Exam No. 8155 failed in several respects to meet test validity standards, particularly those specified in the Guidelines. Whether or not these deficiencies are fatal, they are plainly more substantial than the defects deemed not to defeat validity in prior cases. *Detroit Police Officers Association v. Young,* 446 F. Supp. 979, 990-91, 1007-08 (E.D. Mich. 1978); *Bridgeport Guardians v. Bridgeport Police Department,* 431 F. Supp. 931 (D. Conn. 1977); *cf. Washington v. Davis,* 426 U.S. 229, 248-52 (1976) (Fourteenth Amendment case involving some Title VII concepts). Consequently, assessment of Exam No. 8155 necessarily carries this Court into difficult areas of judging test validity. We must determine, with some care, what the general standards are for judging validity, and how these standards are to be applied in a specific factual situation.

The study of employment testing, although it has necessarily been adopted by the law as a result of Title VII and related statutes, is not primarily a legal subject. It is part of the general field of educational and industrial pychology, and possesses its own methodology, its own body of research, its own experts, and its own terminology. The translation of a technical study such as this into a set of legal principles requires a clear awareness of the limits of both testing and law. It would be entirely inappropriate for the law to ignore what has been learned about employment testing in assessing the validity of these tests. At the same time, the science of testing is not as precise as physics or chemistry, nor its conclusions as provable. While courts should draw upon the findings of experts in the field of testing, they should not hesitate to subject these findings to both the scrutiny of reason and the guidance of Congressional intent.

The need to modify rigid technical conclusions from the field of testing is indicated by the view of certain testing experts, including those who testified for the plaintiffs in this case, that there is no test that can be considered completely valid to select candidates for any but the most rudimentary tasks. If this view guided interpretation of Title VII, then at the current stage of the technology of testing, no test that produces a disparate racial impact could be used for positions such as police officers.

While this position is a conceivable one, it is supported neither by the statutory language nor by judicial precedent. Had Congress felt that testing for virtually all employment was invalid, it would not have made a specific exception to Title VII for the proper use of professionally designed employment tests.[8] Clearly, Congress did not intend that the standard used in interpreting Title VII would reject every test with a disparate racial impact. Nor have the courts permitted this to occur; although they have subjected employment testing to the careful scrutiny required by the statute, they have found a variety of tests to be valid, despite a disparate racial impact. See, *e.g.*, *Sims v. Sheet Metal Workers, Local 65,* 489 F.2d 1023, 1025-26 (6th Cir. 1973); *Detroit Police Officers Association v. Young, supra,* 446 F. Supp. at 1007-08; *Friend v. Leidenger,* 446 F. Supp. 361 (E.D. Va. 1977), *aff'd,* 588 F.2d 61 (4th Cir. 1978); *United States v. South Carolina,* 445 F. Supp. 1094 (D.S.C. 1977), *aff'd,* 434 U.S. 1026 (1978); *Bridgeport Guardians v. Bridgeport Police Department, supra,* 431 F. Supp. at 936-39; *Jackson v. Nassau County Civil Service Commission,* 424 F. Supp. 1162 (E.D.N.Y. 1976); *Buckner v. Goodyear Tire &*

---

8 Title VII states, in relevant part:

> Notwithstanding any other provision of this subchapter, it shall not be an unlawful employment practice . . . for an employer to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin.

42 U.S.C. § 2000e-2(h). *Cf. International Brotherhood of Teamsters, supra,* 431 U.S. at 348-56 (interpreting Title VII's seniority exception, § 2000e-2(h), to give operative effect to exception's full scope).

Rubber Co., 339 F. Supp. 1108, 1113-16 (N.D. Ala. 1972), *aff'd*, 476 F.2d 1287 (5th Cir. 1973).

The danger of too rigid an application of technical testing principles is that tests for all but the most mundane tasks would lack sufficient validity to permit their use. At least that is the risk given the current state of the art of employment testing. This risk can be appreciated by considering the one example even most test critics acknowledge to have substantial, though not complete, validity. This is a typing test given to a group of applicants for jobs as typists.[9] Such a test substantially meets all the criteria suggested by plaintiffs' experts for content validation, but the very success of this test casts doubt on the usefulness of the example. To begin with, typing is a task that readily yields to quantitative measurement. The quality of a typist's job performance depends on two factors, both of which can be captured with precision in numbers: how fast he types, and how many errors he commits. Most jobs involve tasks whose performance can be evaluated only in the more subjective light of judgment. Surely this is true of nearly all the tasks required to be performed by police officers. In addition, there is a more basic problem with the typing test example. Typing is one of the few activities that a test-taker can perform in virtually the same manner as he will be required to perform on the job. That is obviously an ideal testing situation, but it is not one that is frequently available, and such "on-the-job" testing could not possibly be done to select police officers. Yet the force of the typing test example easily leads to one of the conclusions of the District Court in this case: that Exam No. 8155 lacked validity because it measured performance in an artificial classroom setting and did not necessarily indicate who would perform well on the job.[10]

---

9 *See* EEOC, Uniform Employee Selection Guidelines: Interpretation and Clarification (Questions and Answers) Q. 62 1979 [hereinafter cited as EEOC Questions and Answers]; APA Standards, *supra* at 29; *accord, Vulcan Society, supra*, 490 F.2d at 395.

10 In criticizing the questions involving application of the law, the District Court stated: "In a real situation the officer sees activity and must determine rather quickly whether the activity is illegal, with no

Closely related to the question of the proper weight to be given to technical conclusions of testing theory is the question of the proper weight to be given to the EEOC Uniform Guidelines, which are largely based on these technical conclusions. See Guidelines § 5(C). The District Court drew its methodology from the Guidelines, concluding that the City's test was invalid because it failed to satisfy all of the Guidelines. The Supreme Court has relied upon some of the Guidelines in several of the leading cases, see *Albemarle, supra,* 422 U.S. at 431; *Espinoza v. Farah Manufacturing Co.,* 414 U.S. 86, 94 (1973); *Griggs, supra,* 401 U.S. at 433-34, but the Court has not ruled that every deviation from any of the Guidelines automatically results in a violation of Title VII. The Court appears to have applied the Guidelines only to the extent that they are useful, in the particular setting of the case under consideration, for advancing the basic purposes of Title VII. See *Espinoza, supra,* 414 U.S. at 94; *Guardians Association v. Civil Service Commission,* 490 F.2d 400, 403 n.1 (2d Cir. 1973), *United States v. Georgia Power Co.,* 474 F.2d 906, 913 (5th Cir. 1973). To the extent that the Guidelines reflect expert, but non-judicial opinion, they must be applied by courts with the same combination of deference and wariness that characterizes the proper use of expert opinion in general. See *Albemarle, supra,* 422 U.S. at 449 (Blackmun, J., concurring) (Guidelines "have never been subjected to the test of adversary comment. Nor are the theories on which the Guidelines are based beyond dispute.") Thus, the Guidelines should always be considered, but they should not be regarded as conclusive unless reason and statutory interpretation support their conclusions. As this Court has previously stated: "If the EEOC's interpretations go

---

definitional aids before him. He must operate on instinct and experience." 484 F. Supp. at 797. That is true, but it is a criticism of all testing. In any situation, there are generally at least two steps that are necessary to produce the correct behavioral response. The first is to know what to do, and the second is to act accordingly. Clearly the limits of any test, no matter how well designed, is to determine whether the applicant knows or can determine what to do. Only a probationary period can determine if the applicant will act correctly in a real life situation.

beyond congressional intent, the Guidelines must give way." *Guardians Association, supra,* 490 F.2d at 403 n.1.

In addition to their force as the expression of expert opinion, the Guidelines also possess legal force. But here too, it is necessary to keep their limits in mind. The primary purpose of the Guidelines is to indicate the standards that various Federal agencies, such as the EEOC, the Civil Service Commission, and the Department of Justice are to use in enforcing Title VII and related statutes. See Guidelines § 2(A). But the fact that an agency or group of agencies has announced the standards they will use does not convert those standards into mandatory legal rules.

A second legal basis for following the Guidelines is that they represent the "administrative interpretation of the Act by the enforcing agency," and are "entitled to great deference" on that basis. *Griggs, supra,* 401 U.S. at 433-34; see *Albemarle, supra,* 422 U.S. at 431. However, the Court has also recognized that the Guidelines "are not administrative 'regulations' promulgated pursuant to formal procedures established by Congress." *Ibid.* They are entitled to deference, not obedience. See *Espinoza, supra,* 414 U.S. at 94 (1973) (Guideline rule on discrimination against non-citizen "is no doubt entitled to great deference . . ., but that deference must have limits where, as here, application of the guideline would be inconsistent with an obvious congressional intent"). Moreover, the Court in *Griggs* was following the Guidelines only to make the straightforward distinction between general intelligence tests and job-related tests; it is not at all clear that *Griggs* requires observance of all the intricate details of the Guidelines. It might be desirable for all employers to follow the more careful practices required of the Federal Government, but there is no reason to think that Congress intended to impose such practices, in their full rigor, when it enacted Title VII.

With these considerations in mind, we turn to the validity of Exam No. 8155.

IV. The Validity of Exam No. 8155

A. Selecting the Validation Technique

The threshold task in determining the validity of a challenged examination is to select the appropriate method for assessing its job-relatedness. The Guidelines describe three techniques: content validation, construct validation, and criterion-related validation. Guidelines §§ 5(B), 14. The Guidelines specify when each technique is appropriate and also specify the requirements for successfully validating an exam by use of each technique. Defendants have attempted to justify Exam No. 8155 by content validation, a technique appropriate for tests that measure "knowledges, skills or abilities" representative of the "content" of the job. Guidelines § 14(C)(1). Plaintiffs contend that construct validation must be used to assess this exam because, in their view, the exam attempts to measure "constructs," that is, inferences about mental processes or traits, such as "intelligence, aptitude, personality, commonsense, judgment, leadership and spatial ability." *Ibid.*

This content-construct distinction has a significance beyond just selecting the proper technique for validating the exam; it frequently determines who wins the lawsuit. Content validation is generally feasible while construct validation is frequently impossible. Even the Guidelines acknowledge that construct validation requires "an extensive and arduous effort." Guidelines § 14(D)(1). The principal difficulty with construct validation is that it requires a technique that includes a criterion-related study, Guidelines § 14(D)(4)—a demonstration from empirical data that the test successfully predicts job performance.[11] Developing such data is difficult, and tests for which it is required have frequently been declared invalid.[12] As a result,

---

11 Such data may be obtained by studying correlations of test scores of accepted candidates with their subsequent job performances, or correlations of the test scores of present employees with their current job performances.

12 *See United States v. City of Chicago, supra,* 549 F.2d at 430-32; *Douglas v. Hampton,* 512 F.2d 976, 985-86 (D.C. Cir. 1975); *Vulcan Society, supra,* 490 F.2d at 395 & n.10; *Bridgeport Guardians, Inc. v.*

a conclusion that construct validation is required would often decide a case against a test-maker, once a disparate racial impact has been demonstrated.

To determine whether defendants are entitled to use content validation, we examine the Guidelines' criteria for that technique, but we do so bearing in mind our cautionary approach to the Guidelines, previously expressed. The Guidelines specify two basic conditions that must be met before content validation may be used. First, it must appear that what the test attempts to measure is knowledge or an ability, and not a general trait, such as intelligence. Guidelines § 14(C)(1). Second, the test must not measure knowledge or ability that an employee will be expected to learn on the job. *Ibid.*; see also § 5(F). The District Court rejected content validation, concluding both that Exam No. 8155 measures constructs, not abilities, and that, even if what was tested for could be considered abilities, they could be learned in the five-month training program.

In specifying how the selection of validation techniques is to be made, the Guidelines adopt too rigid an approach, one that is inconsistent with Title VII's endorsement of professionally developed tests. Taken literally, the Guidelines would mean that any test for a job that included a training period is almost inevitably doomed: if the attributes the test attempts to

---

Civil Service Commission, 354 F. Supp. 778 (D. Conn), *aff'd in part, rev'd in part*, 482 F.2d 1333 (2d Cir. 1973); *cf.* Schmidt & Hunter, The Future of Criterion-Related Validity, 33 Personnel Psych. 41, 48 (1980) ("criterion related validity studies will frequently, perhaps typically, be technically infeasible").

A rare example of a criterion-related study that was found acceptable is *Washington v. Davis, supra*, 426 U.S. at 249-52, 251 n.17. This was not a Title VII case, however, and the Court's use of Title VII concepts, in dictum, to assess the validity of the test in question under the Fourteenth Amendment does not indicate that the Court was reviewing the test with the stringency that Title VII requires. *See Guardians Ass'n v. Civil Service Commission*, No. 79-7377 (2d Cir. 1980) (slip op. at ____-____). In fact, a less demanding standard was almost certainly being used, as is clear from the comparison between *Davis* and the Court's Title VII decision the preceding term in *Albemarle*.

measure are too general, they are likely to be regarded as constructs, in which event validation is usually too difficult to be successful; if the attributes are fairly specific, they are likely to be appropriate for content validation, but this too will prove unsuccessful because the specific attributes will usually be learned in a training program or on the job.

The origin of this dilemma is not any inherent defect in testing, but rather the Guidelines' definition of "content." This definition makes too sharp a distinction between "content" and "construct," while at the same time blurring the distinction between the two components of "content": knowledge and ability. The knowledge covered by the concept of "content" generally mean factual information. The abilities refer to a person's capacity to carry out a particular function, once the necessary information is supplied. Unless the ability requires virtually no thinking, the "ability" aspect of "content" is not closely related to the "knowledge" aspect of "content"; instead it bears a closer relationship to a "construct." Some researchers regard content tests as nothing more than assessments of particular kinds of constructs, *e.g.*, Tenopyr, Content-Construct Confusion, 30 Personnel Psych. 47 (1977); others regard any ability that is evidenced by observable behavior as sufficiently non-inferential to be considered content, see Ebel, Comments on Some Problems of Employment Testing, 30 Personnel Psych. 55 (1977). See generally Catell, Validity and Reliability: A Proposed More Basic Set of Concepts, 55 J. Ed. Psych. 1 (1964). Whichever view is adopted, it would seem that abilities, at least those that require any thinking, and constructs are simply different segments along a continuum reflecting a person's capacity to perform various categories of tasks. This continuum starts with precise capacities and extends to increasingly abstract ones—from the capacity for filling out forms to the capacity for exercising judgment.

Recognition that abilities and constructs are not entirely distinct leads to a conclusion that a validation technique for purposes of determining Title VII compliance can best be selected by a functional approach that focuses on the nature of the job. The crucial question under Title VII is job related-

ness—whether or not the abilities being tested for are those than can be determined by direct, verifiable observation to be required or desirable for the job. See *Griggs, supra,* 401 U.S. at 431; *Vulcan Society, supra,* 490 F.2d at 394-95; *Chance, supra,* 458 F.2d at 1177. If the job in question involves primarily abilities that are somewhat abstract, content validation should not be rejected simply because these abilities could be categorized as constructs. However, if the test attempts to measure general qualities such as intelligence or common sense, which are no more relevant to the job in question than to any other job, then insistence on the rigorous standards of construct validation is needed. Since tests of this kind are often biased in favor of a person's familiarity with the dominant culture, permitting them to be used without a showing of predictive validity would perpetuate the effects of prior discrimination. But as long as the abilities that the test attempts to measure are no more abstract than necessary, that is, as long as they are the most observable abilities of significance to the particular job in question, content validation should be available. To lessen the risks of perpetuating cultural disadvantages, the degree to which content validation must be demonstrated should increase as the abilities tested for become more abstract.

This functional approach, which adjusts the distinction between content and construct to the nature of the job being tested for, expands the opportunity for both employers and courts to rely on content validation. It also avoids making a threshold choice between content and construct validation based solely on the nature of the quality tested for unrelated to the job, a choice that might make content validation seem inappropriate. To base the content-construct determination on the nature of the job, it is necessary first to analyze the job to see if it requires abilities appropriate for content validation. Instead of choosing between content and construct validation at the outset, as the Guidelines seem to require, employers and courts can start the content validation inquiry and use its results to determine both whether content validation is appropriate and whether it has been achieved. Should the attempted

content validation be found inadequate, the reason may be that this method of validation was not appropriate because of the pertinent job abilities revealed by the job analysis. On the other hand, this approach will sometimes indicate that content validation is appropriate, even though the abilities tested for could be considered constructs.[13]

Just as lessening the severity of the Guidelines' distinction between content and construct reduces the likelihood that a test is invalid because it measures constructs, so sharpening the distinction between knowledge and ability, now obscured by the Guidelines, reduces the problem that the test is invalid because it duplicates the training period, *i.e.*, tests for what will later be learned. Unlike knowledge, some abilities are appropriate for testing confirmed by content validation despite their overlap with post-selection training. A valid measurement of some abilities can select applicants who will ultimately use their training to perform their tasks more effectively or who will more effectively perform similar tasks for which they have not been specifically trained. On the other hand, content validation remains inappropriate for tests that measure knowledge of factual information if that knowledge will be fully acquired in a training program. Approval of such tests, without predictive validation, risks favoring applicants with prior exposure to the information, a course likely to discriminate against a disadvantaged minority. For example, it would be duplicative of the Police Department's training program, and thus invalid, to test applicants for their knowledge of the Department's arrest form. Testing for their ability to fill out the form, however, can be expected to select applicants who

---

13 There will be some tests whose character is sufficiently clear so that the content-construct distinction can be applied at the threshold, without the need to place the test in the context of the job it tests for. A general intelligence test, *see Griggs, supra* (Wonderlic Personnel Test), will almost always need to be assessed by construct validation, since it necessarily measures for an inferred ability, regardless of the context. The much-vaunted typing test, in constrast, can always be regarded as amenable to content validation. However, there are a large number of tests, including virtually all the "second generation" tests for jobs such as the one considered here, that will fall into the middle range.

can be successfully trained to perform well at that task and others like it.

Applying the approach just outlined, we conclude, at least as an initial matter, that content validation may properly be selected as the appropriate technique for assessing Exam No. 8155. The exam tests for three basic abilities (although it purports to test for five): the ability to remember details, the ability to fill out forms, and the ability to apply general principles to specific facts. This third ability is assessed in three contexts: the application of general statements of criminal offenses to the facts of specific events, the application of procedures and standards to the facts of specific situations involving human relations problems. These three basic abilities are not so abstract, on their face, as to preclude content validation, provided subsequent consideration of the job analysis does not demonstrate that important and more concrete abilities necessary for the job were needlessly omitted from those considered for measurement. Though all three abilities involve some inference about mental processes, they are based on observable behaviors and are far less abstract than such traits as intelligence, leadership, or judgment. Moreover, testing for these three abilities sufficiently avoids the objection that the test duplicates the Department's training program. Though all three abilities can be trained to some extent, the test-makers were entitled to select applicants with existing ability so that training would both enhance their abilities and prepare them for other tasks requiring similar talents. The vice of testing for knowledge readily taught in the training program was totally avoided.

B. Assessing the Content Validity of Exam No. 8155

Since content validation appears to be an appropriate method for assessing Exam No. 8155, we proceed to consider whether the use of this method indicates that the exam has sufficient validity to select applicants for the job of police officer. The Guidelines describe various aspects of content validation, but do not neatly list ingredients of an adequate

exam. From our study of the Guidelines, we distill five attributes of an exam with sufficient content validity to be used notwithstanding its disparate racial impact. The first two concern the quality of the test's development: (1) the test-makers must have conducted a suitable job analysis, and (2) they must have used reasonable competence in constructing the test itself.[14] The next three attributes are more in the nature of standards that the test, as produced and used, must be shown to have met. The basic requirement, really the essence of content validation, is (3) that the content of the test must be related to the content of the job. In addition, (4) the content of the test must be representative of the content of the job. Finally, the test must be used with (5) a scoring system that usefully selects from among the applicants those who can better perform the job. We consider each of these five matters in turn.

The Job Analysis

According to the Guidelines, a job analysis involves an assessment "of the important work behavior(s) required for successful performance and their relative importance." § 14(C)(2). The job analysis performed by the City, while somewhat flawed as the District Court pointed out, is nonetheless adequate to meet this standard. As far as the first part of the standard is concerned, the work behaviors involved in being a police officer were identifed by extensive interviewing, and subjected to serious review (Job Analysis, Steps 1 and 2). The District Court found that these work behaviors "were not defined with precision." J. Supp. A. at 18. In fact, the descriptions of the 42 tasks that ultimately appeared on the job analysis list vary considerably in the level of precision. Some are complete and unambiguous, such as "1. Checks the condi-

---

14 These considerations are particularly crucial to an exam validated on content grounds. *See* APA Standards at 29: "Content validity is determined by a set of operations, and one evaluates content validity by the thoroughness and care with which these operations have been conducted."

tion of personal and department equipment such as radio, patrol car, weapons, etc."; "35. Attends training sessions." Others are more open-ended, but do manage to fulfill their function by defining the behaviors associated with the task, such as "3. Performs foot patrol"; "40. Controls various types of crowds." Still others are so vague that they communicate very little real information, such as "10. Interacts with juveniles in non-arrest situations"; "39. Performs duties in hostage situations."

While greater precision might have been achieved, a complete description of the observable tasks associated with being a police officer would be a reworded version of the entire training manual. The Police Department's list of tasks, despite some lapses in specificity, contains a sufficient amount of meaningful information to satisfy the relevant requirement.

The second part of the Guideline standard for a job analysis requires determination of relative importance of the identified work behaviors. The City performed this function by means of an extensively distributed questionnaire, specifying the criteria to be used in ranking the 42 tasks (Job Analysis, Step 3). The process as a whole appears to be reasonably accurate, and neither the plaintiffs nor the District Court raised any serious objection to it.

Having determined the work behaviors and established their relative importance, the City then grouped the 42 tasks into five clusters and asked panels of police officers to identify the knowledges, skills, or abilities necessary to the effective performance of these tasks. (Job Analysis, Steps 4 and 5). This function was implemented in a much less satisfactory manner. Only one of the panels identified the abilities; the other four used the list of five abilities that the first panel had developed. This lessened the value of having five independent panels make this complicated and subjective determination. Moreover, no effort was made to explain the relationship between any of the five abilities and the 42 job tasks from which they were ostensibly derived.[15]

---

15 In some instances the relationship is obvious. Plainly the ability to fill out forms is needed for task 16, "Processes arrests using appropri-

The plaintiffs criticize the required abilities identified by the City for being undefined. But the type of definition suggested by the Guidelines—one that describes the abilities in terms of "observable behaviors and outcomes," Guidelines § 15(C)(3)—seems repetitive, since the work behaviors are already defined in this way. The five identified abilities, with the possible exception of "human relations, including communication techniques," are comprehensible enough. Their appropriateness for measurement would have been considerably clearer, however, if each panel had explained which tasks required which abilities. While the Guidelines may be unnecessarily stringent in regarding the identification of this relationship of ability to task as "essential," see *ibid.*, such identification does go far toward eliminating the ambiguities that are otherwise inherent in generalized descriptions of abilities. Only if the relationship of abilities to tasks is clearly set forth can there be confidence that the pertinent abilities have been selected for measurement.

The Test Construction Process

With a job analysis of questionable sufficiency, the City then proceeded to the test construction stage. As an initial matter, we note that Exam No. 8155 was developed "in-house," by staff members of New York City's Police Department and Department of Personnel; there was little input from any outside source, and no participation by anyone specializing in test preparation. Of course, the law should not be designed to subsidize specialists. But employment testing is a task of sufficient difficulty to suggest that an employer dispenses with expert assistance at his peril. Certainly, the decision to forgo such assistance should require a Court to give the resulting test careful scrutiny. See *Kirkland, supra,* 520 F.2d at 425-26; *Vulcan Society, supra,* 490 F.2d at 395-96.

---

ate police department forms and notifications." But it is not evident, for example, why any of the five listed abilities are critical to task 32, "Searches for lost children, runaways, etc." or task 19, "Guards and transports prisoners."

While the determination of how many questions should be included for each identified ability was made by a fairly careful numerical analysis (Test Construction, Step 1), the process of writing the questions themselves was rather haphazard. The questions were initially framed by police officers, who may have had expertise in identifying tasks involved in their job but were amateurs in the art of test construction. In addition, the officers did not have access to the job analysis material during much of the process. Finally, the questions, although they were reviewed, were not tested on a sample population. To be sure, a complete determination of the questions' accuracy in measuring the identified abilities would be equivalent in its complexity to a criterion-related study. But the City did not even perform the minimal sample testing to ensure that the questions were comprehensible and unambiguous.

Not surprisingly, the test construction process did not fully succeed in meeting even its own goal of testing for all the identified abilities. As previously indicated, Exam No. 8155 does appear to test for the three identified abilities of remembering details, filling out forms, and applying general principles to specific facts. However, the fourth identified ability, human relations skill, proved more troublesome. In deciding how to test for this ability, the City faced a dilemma inherent in testing for all but the most mundane jobs. To be fully representative of the job, a test should measure all the significant abilities needed for successful job performance, yet some abilities, especially in jobs of any complexity, are far along the construct end of the content-construct continuum where successful validation is difficult. If a test tries to be representative and measure all significant abilities, including those that are clearly constructs, it risks the use of inadequate assessment devices, because the rigorous standard for construct validation will rarely be met. On the other hand if the test-makers acknowledge the difficulty of satisfactorily measuring constructs and test only for those abilities that are appropriate for content validation, they encounter the objection that the test is not sufficiently representative of the job.

Recognizing the difficulty of construct validation, yet reluctant to omit assessment of an important characteristic of successful job performance, the City attempted to resolve the dilemma by treating human relations skill as an ability suitable for content validation and devoting 30 questions, nearly one-third of the exam, to an effort to assess this ability. Mindful of an important requirement of content validity, the City carefully avoided rewarding a test-taker's prior knowledge and, instead, supplied in the test itself all the information necessary to select the correct answers to the human relations questions. Included before each group of questions was a set of appropriate standards—essentially "do's" and "don'ts"—for handling a particular type of human relations matter. But supplying this guidance rendered the 30 questions primarily a further assessment of a candidate's ability to apply written standards to specific fact situations, and only slightly a measure of his talent for human relations. Anyone with minimal analytic ability needed to apply the standards to the various fact situations could select the one correct answer, even if his intuitive reaction to a human relations problem might be woefully inadequate.

Assessing human relations skill will always be a difficult enterprise, but the deficiency of the City's attempt does not mean that a context validation approach is necessarily impermissible nor impossible to achieve. As indicated above, at least within the middle range of the content-construct continuum, the distinction between context and construct should be determined functionally, in relation to the job. If the quality measured is not unduly abstract, and if it constitutes a significant aspect of the job, content validation of the test component used to measure that quality should be permitted. But that component must be designed in an extremely careful way. Test-makers will be well advised to obtain highly qualified assistance in constructing this portion of an exam.

One desirable approach would be to confront applicants with simulated real life situations and assess the appropriateness of their volunteered responses. See *Firefighters Institute for Racial Equality v. City of St. Louis,* ____ F.2d ____ (8th

Cir. 1980). That technique is normally too costly for large numbers of applicants, but might have usefulness as a testing device to be used toward the end of the overall selection procedure, after an initially large group of applicants has been narrowed down by the results of a written exam and a background check. If the test component is limited to traditional pencil and paper methods, it may be preferable to forgo any pretense of being able to make fine differentiation among candidates' human relations skills and instead adopt a pass/fail approach, rejecting those whose demonstrably inappropriate responses to human relations questions mark them as unsuitable for police work. Another possibility is to recognize that questions in this area for which only one answer is correct are likely to be too easy, as were most of the questions on Exam No. 8155, and therefore of little use in making selections from among applicants. Instead questions can be designed for which some answers are appropriate responses and others are inappropriate. As feasible techniques in this area evolve, employers will be expected to use them.

With these strengths and weaknesses of the job analysis and the test construction in mind, we now consider how well the test, as constructed and used, met the basic requirements of content validity.

The Direct Relationship Requirement

The central requirement of Title VII, relationship of test content to job content, was sufficiently satisfied by Exam No. 8155. The job analysis procedure provides adequate assurance that the identified tasks are in fact the tasks that a police officer performs. While the procedure for identifying the abilities required for those tasks was less satisfactory, the three abilities that were actually tested for appear adequately related to most of the identified tasks. The list of tasks confirms one's intuitive assumption that police officers are required to fill out forms (see, *e.g.*, "16. Processes arrests using appropriate police department forms and notifications"), to remember facts (see, *e.g.*, "18. Gives testimony in court (oral and written)"), and to

apply general principles to specific fact situations (see, *e.g.*, "26. Executes warrants.").

Moreover, these abilities are among the most concrete ones that can be derived from the list; they are certainly more concrete than human relations skills, which the test purported to measure, but did not. Two of the abilities tested for, filling out forms and remembering facts, are as specifically stated as they could be without resort to trivial distinctions about particular kinds of forms and facts. The ability to apply general standards is somewhat more problematical, since it is a relatively abstract skill that is relevant to many jobs. However, if there is any job for which ability in applying and following rules is an especially important requirement, it is the job of a law enforcement officer.[16]

The Representativeness Requirement

The second requirement established by the Guidelines is that the test must be a "representative sample of the content of the job." As presented by the Guidelines, this representativeness requirement has two different meanings. The first is that the content of the test must be representative of the content of the job; the second is that the procedure, or methodology, of the test must be similar to the procedures required by the job itself. The Guidelines express this dual requirement in the following somewhat inscrutable language: "For any selection procedure measuring a knowledge, skill, or ability the user should show that (a) the selection procedure *measures and is* a representative sample of that knowledge, skill, or ability. . . ." Guidelines § 14(C)(4) (emphasis added).

---

16 The fact that the factual subject matter of the exam questions (as opposed to their purpose in measuring abilities) was related to the subject matter of the job is not a major indicator of the test's validity, since the test measured abilities, not knowledge. But it does suggest that the test has avoided the dangers inherent in using irrelevant factual material. Such material could skew the test for ability in directions unrelated to the job, a phenomenon that even the best designed test might not be able to avoid. The present test avoids that problem by ensuring that any distorting effects resulting from the subject matter of the questions are themselves job-related.

Both aspects of the representativeness requirement, if interpreted rigorously, would once again foreclose any possibility of constructing a valid test. The United States, as *amicus,* argues that the requirement that the content of the exam be representative means that all the knowledges, skills, or abilities required for the job be tested for, each in its proper proportion. This is not even theoretically possible, since some of the required capacities cannot be tested for in any valid manner. Even if they could be, the task of identifying every capacity and determining its appropriate proportion is a practical impossibility.

It is similarly impossible for the procedures of the test to be truly representative of the actual job procedures. Tests, by their nature, are a controlled, simplified version of the job activities, not the activities themselves. As a practical matter, virtually any realistic test, except one that directly measures a physical skill, like lifting 50-pound sacks, is likely to be a pencil and paper activity, quite different from the job it tests for. An elaborate effort to simulate the actual work setting would be beyond the resources of most employers, and perhaps beyond the capacities of even the most professional test-makers.

More reasonable interpretations of the representativeness requirement are appropriate in light of Title VII's basic purposes. The reason for a requirement that the *content* of the exam be representative is to prevent either the use of some minor aspect of the job as the basis for the selection procedure or the needless elimination of some significant part of the job's requirements from the selection process entirely; this adds a quantitative element to the qualitative requirement—that the content of the test be related to the content of the job. Thus, it is reasonable to insist that the test measure important aspects of the job, at least those for which appropriate measurement is feasible, but not that it measure all aspects, regardless of significance, in their exact proportions. The reason for a requirement that the test's *procedure* be representative is to prevent distorting effects that go beyond the inherent distortions present in any measuring instrument. For example, al-

though all pencil and paper tests are dependent on reading, even if many aspects of the job are not, the reading level of the test should not be pointlessly high. Similarly, the instructions should not be overly complex, and the exam should not place candidates under excessive time pressure unless such time pressure is an identifiable aspect of the job.

Exam No. 8155 meets these representativeness requirements to an adequate degree. While it did not test for all the skills involved in being a police officer nor adequately test for the human relations skill that the job analysis identified as important, the ones it did measure—memory, the ability to fill out forms, and the ability to apply rules to factual situations—are all significant aspects of entry-level police work. To be sure, this conclusion would have been easier to reach if the City had spelled out the relationship between the abilities that were tested for and the job behaviors that had been identified. But the relationship is sufficiently apparent to indicate that the City was not seizing on minor aspects of the police officer's job as the basis for selection of candidates. The inadequate assessment of human relations skill lessens the representativeness of the exam and consequently lessens its degree of content validity, but this deficiency is not fatal, especially in light of the difficulty of assessing such an abstract ability. Though human relations skill was deemed so important as to warrant 30 of the exam's 100 questions, the City could just as plausibly have concluded that equally important for the job of policing are such other abstract qualities as common sense, leadership potential, sound judgment, or ability to resist provocation. When a police exam inadequately tests for any of these abstract abilities, it simply recognizes the limits of the art of testing. Indeed, the more a test concerns itself with relatively concrete abilities identified as necessary for successful job performance, the more likely it is to achieve a sound basis for assessment of applicants.

Similarly, the procedure that the test employed was not needlessly unrepresentative of the job itself. In electing to use a pencil and paper test, the City did not forgo any readily available and realistically feasible alternative procedure that

would have been more representative of the job. Moreover, the risks of using a written test were substantially minimized. The reading level necessary to understand the questions was in some cases equal to, but generally well below, the training materials used in the Police Academy. The instructions were clear enough, and employed an ordinary four answer multiple choice format, perhaps the most familiar standardized test technique. In addition, ample time was allowed for taking the exam, thereby avoiding an unnecessarily pressured situation.

Thus the exam is adequately related to the content of the policy officer's job, and adequately representative. The combined effect of this assessment might support a conclusion that the exam as a whole has content validity, though it would be a close question whether a test with the disparate racial impact of this one can be validated when its development departs in some significant respects even from reasonably attainable requirements of the Guidelines. However, even if the construction of the exam passes muster, the way in which it was used to distinguish among candidates seriously departs from the third requirement for content validity and defeats any claim of validity for a testing process that produces disparate racial results.

The Scoring Requirement

Essentially, the City used the results of the exam to compile a rank-ordering of all the applicants, and then selected a passing score sufficient to generate the required number of potential trainees. Neither the rank-ordering nor the passing score conforms to even the most minimal standards for these two devices.

Rank-Ordering. The Guidelines provide that rank-ordering should be used only if it can be shown that "a higher score . . . is likely to result in better job performance." Guidelines § 14(C)(9). This requirement is reasonable and consistent with Title VII's provision that the "results" of a test may not be "used to discriminate." 42 U.S.C. § 2000e-2(h). If test scores do not vary directly with job performance, ranking the candi-

dates on the basis of their scores will not select better employees. It is possible to read the Guidelines' standard for rank-ordering as if the required relationship between better scores and better job performance had to be demonstrated by a criterion-related study. However, the EEOC's interpretation of the Guidelines disclaims such a high standard. The relationship between higher scores and better job performance may permissibly rest on an inference, but where, as here, the test scores reveal a disparate racial impact, and that disparity is greater at high passing scores than at low passing scores, the appropriateness of inferring that higher scores closely correlate with better job performance must be closely scrutinized.

This close scrutiny is required because rank-ordering makes such a refined use of the test's basic power to distinguish between those who are qualified to perform the job and those who are not. If a test is content valid, it may be reasonable to infer that the test scores make some useful gross distinctions between candidates. Candidates with high scores may well be expected to perform the job better than candidates with low scores. See Science Research Associates, *Validation: Procedures and Results* (1972) (use of criterion "tails" identifying best and worst candidates more justifiable than continuous rating). And it may even be that within some range of scores, some incremental improvements in scores show some positive correlation with improvements in job performance. But neither of these propositions provides confidence for inferring that one-point increments among those who took Exam No. 8155 are a valid basis for making job-related hiring decisions, especially in the range of scores between 94 and 100. The reason such a precise inference cannot be so readily drawn is that content validity is not an all or nothing matter; it comes in degrees. A test may have enough validity for making gross distinctions between those qualified and unqualified for a job, yet may be totally inadequate to yield passing grades that show positive correlation with job performance.

Overlooking this point, the City earnestly contends that if the appropriate abilities were tested for, it makes eminent sense to select candidates strictly on the basis of ranked scores, even

to the extent of concluding that a candidate scoring 98 will perform better as a police officer than a candidate scoring 97. The frequency with which such one-point differentials are used for important decisions in our society, both in academic assessment and civil service employment, should not obscure their equally frequent lack of demonstrated significance. Rank-ordering satisfies a felt need for objectivity, but it does not necessarily select better job performers. In some circumstances the virtues of objectivity may justify the inherent artificiality of the substantively deficient distinctions being made. But when test scores have a disparate racial impact, an employer violates Title VII if he uses them in ways that lack significant relationship to job performance.

Permissible use of rank-ordering requires a demonstration of such substantial test validity that it is reasonable to expect one- or two-point differences in scores to reflect differences in job performance. Our prior conclusion that the test itself may have had enough validity to be used does not, therefore, lead to approval of using its results for rank-ordered selections. On the contrary, the defects we noted in the job analysis and the test construction are substantial enough to preclude an inference that passing scores will correlate with job performance closely enough to justify rank-ordered selections. While we do not criticize the City's efforts as extensively as did the District Court, we agree that the identification of pertinent abilities, the demonstration of their relationship to the job tasks, and the process of developing the questions, were flawed.

These shortcomings take on added significance when it is recognized that the test just barely satisfied even our lenient construction of the Guidelines requirement of procedural representativeness. As the EEOC has advised, it is "easier" to make the inference of a relationship between higher scores and better job performance "[t]he more closely and completely the selection procedure approximates the important work behaviors." EEOC Questions and Answers, *supra,* Q. 62. Unlike the District Court, we are not willing to reject any use of a police exam simply because the pencil and paper procedure of the test is not a close approximation of the job. Nor are we willing to

preclude rank-ordering because a pencil and paper procedure was used. Given the current state of the art in employment testing, we think it would be unrealistic to condemn pencil and paper tests. Alternative procedures have not been shown to be readily available within the limitations of time and resources confronting most employers. Nevertheless, we cannot ignore the Guidelines' criticism of assessing ability to perform complex tasks by a test procedure so different from the work setting. When the selection procedure does not closely approximate the important job tasks, it becomes especially important to insist upon a strong showing that other aspects of content validity have been demonstrated. And that demonstration must be very substantial when a test procedure that does not closely approximate the job is sought to reflect the fine gradations required for rank-ordering. In short, while we might not agree with the District Judge that the defects in the test preclude a finding of sufficient content validity to permit its use, we agree that content validity has not been shown to the extent necessary for rank-ordering.

In addition to inadequate demonstration of validity, the test may not be used for rank-ordered selections because of the total absence of any evidence that the exam possessed another vital feature—reliability, that is, the extent to which the exam would produce consistent results if applicants repeatedly took it or similar tests. Of course, there is no expectation that applicants *will* take any given test more than once. But if an exam lacks reliability to such an extent that results *would* be significantly inconsistent if the same applicants were to take it again, that is an important indication that the test is not especially useful in measuring their abilities. Although not explicitly mentioned in the Guidelines reliability is prominently identified in the APA Standards (to which the Guidelines refer in § 5(C)) to be as basic for evaluating an exam as validity itself. See APA Standards, *supra* as 48-55. Like content validity, reliability is not an all or nothing matter. It too comes in degrees. What is required is not perfect reliability, but rather a sufficient degree of reliability to justify the use of being made of the test results. Without some substantial demonstration of

reliability it is wholly unwarranted to make hiring decisions, with a disparate racial impact, for thousands of applicants that turn on one-point distinctions among their passing grades.

Two aspects of reliability deserve consideration in assessing the use of rank-ordering. The first is the quality of the exam questions. The more skillfully they have been formulated, the more likely it is that results on one question will correlate with results on the other questions and that successive test scores would be consistent. This will avoid the tendency of scores to vary because of extraneous factors such as test administration. Whether this aspect of reliability has been achieved to an extent sufficient to justify rank-ordering need not be left to general consideration of the quality of the test construction process. A basic demonstration of this aspect of reliability can easily be made by the test-maker, before the test is administered to job applicants. The test-maker can pre-test his exam by giving it twice to a sample of persons generally approximating the characteristics of the population where the test is expected to be used for employment selection. To avoid distortion due to recollection, the test given at a later date to such a sample can use similar but not identical questions. Another somewhat useful indicator of reliability is a technique known as a split-half correlation—dividing each component of the test into equal halves and observing how consistent were an individual's scores on each half.[17] This technique can also be used in the process of pre-testing the exam, before it is administered to job applicants. The technique also is easily used on actual test results to provide some minimal evidence of reliability. In this case the City offered no evidence to demonstrate the quality of the questions used in Exam No. 8155.

The second aspect of reliability concerns what testing experts call the error of measurement. See generally, H. Gulliksen, The Theory of Mental Test (1950). This is a statistical phenomenon

---

17 *See* APA Standards, *supra* at 48-50; Kuder & Richardson, The Theory of Estimation of Test Reliability in Principles of Educational and Psychological Measurement 95 (W. Mehrens & R. Ebel, eds. 1967); Rulon, A Simplified Procedure for Determining the Reliability of a Test by Split Halves, in *id.* at 104.

indicating the degree to which scores on successive tests will be subject to inevitable random variation, no matter how carefully the test-makers have eliminated or at least lessened the effects of extraneous factors within their control. The error of measurement can be calculated by use of the standard deviation concept. For any test, regardless of how carefully it was prepared, statistical analysis, based on the normal distribution curve, shows that there is 68% probability that successive scores would fall within a range of one standard deviation from an actual score and a 95% probability (generally a satisfactory confidence level) that successive scores would fall within a range of two standard deviations from the actual score. It is also possible to estimate, again for any test, how many raw score points above and below the applicant's actual score are within the range of one or more standard deviations. This calculation, as explained in the margin,[18] depends upon the applicant's score and the number of items on the test. Thus, though the test-maker can never eliminate the error of measurement, he can minimize its effect for all scores by increasing the number of questions.

---

18 This effect can be demonstrated more precisely in psychometric terms, through the use of the standard error concept. The standard error is the raw score variance corresponding to a single standard deviation of the scores that would be obtained by a test-taker on successive, equivalent tests. It can be approximated by the quantity

$$\sqrt{\frac{t(n-t)}{n}}$$

where t is the test-taker's score and n is the number of items on the test. This formula is derived from the general formula for a standard deviation. *See* Lord, *Do Tests of the Same Length Have the Same Standard Error of Measurement?*, in *Principles of Educational and Psychological Measurement* 192 (W. Mehrens & R. Ebel, eds. 1967). The formula is an approximation because a particular applicant's error of measurement as defined by Lord, *supra*, is a function of his "true" score, that is, the score he would have obtained had no error been present. However, as the number of items on the test increases, the observed score will approach the true score, so that the approximation is permissible. In substituting observed score for actual score Lord introduces the refinement of reducing the denominator of his function by one to eliminate sampling bias, but this has an insignificant effect when the numbers are rounded off to the extent that they are in this discussion.

The inevitable error of measurement for a test consisting of 100 items, like Exam No. 8155, has significance in assessing the use of rank-ordering. At the passing score of 94, one standard deviation is equivalent to a range between 2.4 points above and below 94. The range narrows as actual scores approach 100. At 97, for example, the range is plus or minus 1.7. Thus, to have 95% confidence that an applicant's grade has statistical reliability, grades within two standard deviations of his grade should theoretically be treated as equivalent to his grade, for in fact there is a 95% likelihood that each applicant at each grade would score within such a range on successive takings of equivalent tests. This means that the range in which a satisfactory confidence level is achieved for an applicant who scores 94 lies between 89 and 99, and even for one who scores 97, the range extends from 94 to 100. Care must be taken not to over-emphasize the significance of the error of measurement. Though grounded on sound principles of statistics, it remains an estimate, and it need not prevent the usual use of test scores that do not have a disparate racial impact. At a minimum, however, it should serve to illustrate the risks of making hiring decisions turn on one-point increments at scores where even a single standard deviation covers a raw score range greater than one point.

The most serious implication of error of measurement for Exam No. 8155 arises from the extraordinary extent to which his test scores were closely bunched.[19] Each score from 94 to 97 was achieved by over 2,000 applicants.[20] If the test questions

---

19 The reason this bunching occurred was that the exam was too easy. An exam that was too difficult might have had the same effect, except that the bunching would have occurred at the lower end of the scale. Neither excessive easiness nor excessive difficulty is necessarily fatal, but each magnifies effects that may make scoring arrangements unjustified.

20 Set forth below are the total number of applicants who achieved each score from 110 to 70 and the number of White and minority (Black and Hispanic surnamed) applicants at each of these scores. The White and minority figures do not always equal the total because some applicants were members of other minority groups and some applicants were not identified.

(Footnote continued)

had sufficient differentiating power to produce a somewhat even distribution of scores, or at least to avoid excessive bunching among the high scores, the error of measurement would not have affected the ultimate selection of such a

Score	*Total Applicants*	*White*	*Minority*
110	3	1	1
109	6	4	2
108	3	1	0
107	9	5	1
106	13	5	5
105	36	19	6
104	90	59	16
103	102	64	23
102	95	53	17
101	125	60	38
100	823	565	96
99	1570	1067	177
98	1845	1372	223
97	2238	1562	282
96	2311	1516	390
95	2255	1434	428
94	2124	1307	425
93	2024	1195	504
92	2017	1174	504
91	1772	1009	524
90	1675	913	529
89	1498	774	541
88	1359	678	485
87	1171	569	452
86	1077	500	452
85	989	468	402
84	870	395	371
83	813	358	362
82	722	313	330
81	608	246	298
80	558	214	287
79	497	191	245
78	449	163	222
77	409	157	197
76	363	130	192
75	323	117	156
74	342	121	182
73	265	89	141
72	247	87	138
71	213	89	102
70	198	65	112

significant portion of the applicants. But when 8,928 applicants, two-thirds of all who passed, are bunched between 94 and 97, the error of measurement makes the use of rank-ordering an extremely unreliable basis for hiring decisions.

If test scores produce disparate racial results, an employer who wants to use rank-ordering of the scores for hiring decisions faces a substantial task in demonstrating that rank-ordering is sufficiently justified to be used. But the task is by no means impossible. Even without resorting to a criterion-related study, the test-maker still has several ways to increase the justification for rank-ordering sufficiently to use it. First, he can conduct a job analysis and construct the test with a high degree of adherence to Guideline requirements. That would produce a much stronger showing of content validation than the City was able to demonstrate in this case. Even content validity sufficient for rank-ordering does not require literal compliance with every aspect of the Guidelines. But there must be a substantial demonstration of job relatedness and representativeness to show a sound basis for making rank-ordering hiring decisions. Second, the test-maker can achieve an adequate degree of reliability by careful design of the exam so that the questions will yield a satisfactory degree of consistent results. To guard against inconsistency based on extraneous factors, the test-maker can pre-test the exam by successive applications to an appropriate sample or at least analyze the results of split-half correlations. Inconsistencies revealed by these techniques can be lessened by redesign of needlessly unreliable questions or components of the exam. To reduce inconsistency based on random variation, the number of questions can be increased. Of course, the size of an exam must observe realistic limits of cost and time of administration, but in the case of Exam No. 8155, using 200 instead of 100 questions would have significantly increased reliability. Because some error of measurement is inevitable, even an increase in the number of questions will not eliminate all random variation. However, the effect of such random variation can be reduced by using questions that are shown to have significant

differentiating power, so that scores are not bunched at the high end of the scale.[21]

Alternatively, the employer can acknowledge his inability to justify rank-ordering and resort to random selection from within either the entire group that achieves a properly determined passing score, or some segment of the passing group shown to be appropriate. The City itself, perhaps unwittingly, has acknowledged the reasonableness of this second alternative. Since each of the scores between 94 and 97 were achieved by more than 2,000 candidates, and since each training class can accommodate slightly more than 400 candidates, the test scores provide no basis for selecting from among candidates at each of these scoring levels. At oral argument, the City acknowledged that random selection would be used; for example, if all candidates scoring 98 or above have been selected, and 400 academy trainees are needed from the 2,000 candidates scoring 97, a random drawing from among all 2,000 would be used.[22] Thus, even the City recognizes that when the test scores afford no job-related basis for making selections from within a group that passed the test, random selection is appropriate.

We do not conclude that Title VII requires random selection from among those who pass a content valid test. In some instances rank-ordering may be shown to be justified. But where it is not, random selection from within a group validly determined to have passed a content valid exam is simply an

---

21 The differentiating power of a question can be easily determined by means of an item analysis. The City could have tested each question on a sample population to determine its power to distinguish between different levels of ability. The most simple item analysis would have quickly revealed that this exam would produce a large number of closely bunched high scores. *See* Englehart, A Comparison of Several Item Discrimination Indices, reprinted in Principles of Educational and Psychological Measurement 387 (W. Mehrens & R. Ebel eds. 1967); Findley, A Rationale for Evaluation of Item Discrimination Statistics, reprinted in *id.* at 381.

22 Actually, in this example, presumably 1,200 names would be drawn to yield the anticipated 400 candidates who would complete the post-examination steps of the hiring process.

available option. See *Association Against Discrimination in Employment v. City of Bridgeport,* 594 F.2d 306, 313 n.19 (2d Cir. 1979). The City may prefer not to use it. However that may be, the City cannot use rank-ordering not shown to be job-related when test scores produce a disparate racial impact. Nor can the City justify the use of rank-ordering by reliance on what it contends are requirements of state law. See N.Y. Const. art. 5, § 6; N.Y. Civil Service Law § 61(1). Title VII explicitly relieves employers from any duty to observe a state hiring provision "which purports to require or permit" any discriminatory employment practice. 42 U.S.C. § 2000e-7 (1976).

If rank-ordering were the only unjustified use of test scores, it would be possible to limit a Title VII remedy to the elimination of this device. In other words, if an exam has adequate content validity and a passing score has been adequately determined, the employer could still limit selections to those within the group that passed, provided only that he abandons rank-ordered choices. But in this case, the impermissible use of the test scores extends beyond rank-ordering to the setting of the cutoff score.

Cutoff Score. The Guidelines state that a cutoff score "should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force." Guidelines § 5(H). This also makes sense. No matter how valid the exam, it is the cutoff score that ultimately determines whether a person passes or fails. A cutoff score unrelated to job performance may well lead to the rejection of applicants who were fully capable of performing the job. When a cutoff score unrelated to job performance produces disparate racial results, Title VII is violated. See *Association Against Discrimination, supra,* 594 F.2d at 312-33; *Bridgeport Guardians, Inc. v. Bridgeport Civil Service Commission,* 482 F.2d 1333, 1338 (2d Cir. 1973). Consequently, there should generally be some independent basis for choosing the cutoff. As with rank-ordering, a criterion-related study is not necessarily required; the employer might establish a valid cutoff score by using a professional estimate of the requisite ability levels, or, at the very least, by analyzing the test results to locate a

logical "break-point" in the distribution of scores. The City offered no such basis in this case. It merely chose as many candidates as it needed, and then set the cutoff score so that the remaining candidates would fail.

If it had been shown that the exam measures ability with sufficient differentiating power to justify rank-ordering, it would have been valid to set the cutoff score at the point where rank-ordering filled the City's needs. The justification would be that each incremental change in score represents an incremental change in job-related ability, so that, for any given cutoff (even one determined solely by hiring needs), those who passed would likely perform the job better than those who failed. But the City can make no such claim, since it never established a valid basis for rank-ordering.

Indeed, the problems of both validity and reliability, which prevent the justified use of rank-ordering, also cast serious doubt on the justification for the cutoff score of 94. Of all these problems, the unreliability attributable to the error of measurement has special significance for the cutoff score. As previously noted, the error of measurement had an especially extensive impact on the applicants because of the bunching of scores at the high end. The bunching occurred not only at passing scores from 94 to 97, but also at failing scores of 92 and 93, each of which was also achieved by more than 2,000 applicants. Scores within a range of two points above and below the passing grade of 94 were achieved by 10,731 applicants, 29% of the total. Had the scores been evenly distributed, 1,800 applicants, only 5%, would have fallen within this range. Selecting a cutoff score in the middle of the range in which the test scores were closely bunched meant that the inevitable error measurement led to a much higher number of mistaken passes and failures than would otherwise have occurred.[23] Perhaps an even distribution of scores cannot be

---

23 This comparison can be expressed mathematically by considering the statistical probability that a person who passed with a score of 94 or failed with a score of 93 would have achieved that same score on successive exams. If one were to administer successive exams and average the results, the test-taker would have to achieve an average

readily achieved, but the impact of the error of measurement could have been held to acceptable limits if a cutoff score had been selected within some range where scores were not closely bunched. This does not mean that every person who fails a test by a single point necessarily has a claim for legal redress. A cutoff score, properly selected, is not impermissible simply because there will always be some error of measurement associated with it. But when an exam produces disparate racial

---

score no more than one-half point above or below his original score to be regarded as achieving the same score. For a test-taker who scored 93 or 94 on Exam No. 8155, the standard error is approximately 2.4. This means that 2.4 points above or below 93 or 94 represents one standard deviation. Similarly, one-half a point above or below these scores represents .5/2.4 or .21 standard deviations. Assuming that the normal distribution curve applies, and ignoring the effect of bonus points, it can be derived from the table of normal curve distributions that .21 standard deviations represent a 17% level of certainty. In other words about 17% of the test takers who scored 93 or 94 would achieve that same score, on the average, if they had taken successive equivalent exams. Of the remainder, we can assume that half would have scored higher and half would have scored lower. Since 94 was passing, these figures mean that 41.5% of those who scored 93 (41.5% being half the remainder, which is 100%—17%, or 83%), would have achieved an average score of 94 on a series of successive equivalent tests, and thereby passed the exam, while 41.5% of those who scored 94 would have failed.

If the test-takers had been evenly distributed by score, which is ideal, or at least approximated an even distribution in the cutoff region, which is generally feasible, 729 persons would have scored 93 or 94, and 41.5% of these, or 300, would have been incorrectly placed. In fact, some 4,148 achieved these two scores, so that at least 1,721 of the test-takers were incorrectly placed, just counting those who achieved these two sets of scores. As one moves away from the cutoff, the percentage of incorrect placements becomes much less. Of those who scored 99, for example, virtually none would have failed on successive tests, since the standard error for a score of 99 on a test with 100 questions is about 1, and 93, the highest failing grade, is thus more than three standard deviations below 99. Similarly, of those who scored 85, only one half of one percent would have passed on successive tests, since the standard error for a score of 85 is 3.6, which is 2.5 standard deviations below 94. For these scores, the test had a very low error of measurement. But relatively few of the test-takers in Exam No. 8155 scored 85 or 99. In contrast, each score in the 92 to 97 range, where the error in measurement was greatest, was achieved by over 2,000 test-takers. And the selected cutoff score was right in the middle of that range.

results, a cutoff score requires adequate justification and cannot be used at a point where its unreliability has such an extensive impact as occurred in this case.

Primarily on the basis of Exam No. 8155's improper use of rank-ordering, and of the cutoff score, we affirm the conclusion of the District Court that the exam as used was invalid. Since we agree with the District Court that the exam had a significant disparate racial impact, we hold that the City's use of the exam violated Title VII.

V. Relief

The fashioning of relief in employment discrimination cases is always a sensitive matter, especially in cases like this one where the District Court endeavors to order some form of affirmative action, including the use of a quota. Our task of determining whether the District Court's remedy conforms to prevailing standards for Title VII relief has been made somewhat more difficult than usual because the precise effect of the Court's order is not clear and, in some respects, the order is not adequately supported by necessary findings or sufficient evidence.

The District Court's order is set out in full in the margin.[24] It deals with several topics including the use of Exam No. 8155,

---

24 Upon consideration of the evidence presented at the liability and relief stages of this case, and consideration of the briefs and oral arguments of the parties and *amicus curiae* United States of America and Policewomen's Endowment Association of New York City, Inc., and entry of findings of fact and conclusions of law, it is hereby ORDERED:

1. Defendants, their officers, officials, agents, employees, successors, and all persons in active concert or participation with them or any of them are hereby permanently enjoined from engaging in any act or practice with respect to the selection of candidates for appointment to and training for the position in the New York City Police Department of entry-level police officer, which act or practice has the purpose or effect of discriminating against such persons because of race or national origin.

(Footnote continued)

the development and approval of a new selection procedure, hiring in the interim until a new selection procedure receives

---

2. Defendants are hereby permanently enjoined from using Examination 8155 in any manner, except as specifically provided in paragraph 5 of this order.

3. The defendants shall seek to achieve as a long-term goal black and hispanic (hereinafter "minority") representation in the sworn ranks of the Police Department comparable to that of the minority composition of the labor force in the relevant hiring area. As of 1978, the labor force of the relevant hiring area was at least 30% black and hispanic.

4. To achieve the long-term goal set forth in paragraph 3, *supra,* the defendants shall as an interim goal appoint 50% of their entry level police officers from among qualified black and hispanic applicants. The interim hiring goal for minorities shall remain in effect until the minority representation in the sworn ranks of the Police Department is at least equal to the percentage of minorities in the labor force of the relevant hiring area as described in paragraph 3, *supra,* or until this court has found, after a hearing, that all proposed selection procedures for police officer positions have been validated in accordance with the *Uniform Guidelines on Employee Selection Procedures,* 28 C.F.R. § 50.14, 29 C.F.R. § 1607, effective September 25, 1978 ("Uniform Guidelines"), and that no further interim goals are appropriate. Nothing herein shall preclude plaintiffs from advocating the continuance of the interim goals on the basis that the continuing effects of past discrimination have not been eliminated.

5. To satisfy the goals set forth above in paragraphs 3 and 4, defendants may use an eligibility list derived from Examination No. 8155 as the pool from which it selects police officers. At such time as that eligibility list or any future eligibility list for the position of police officer does not contain sufficient minority candidates to meet the interim goals set out in paragraph 4, *supra,* the City shall take whatever steps are necessary to achieve the interim hiring goals.

6. Defendants shall make reasonable efforts to develop within a reasonable period of time a procedure for the selection of candidates for the entry level position of police officer which shall be lawful and validated in accordance with the Uniform Guidelines or successor guidelines similarly promulgated, and which is consistent with generally accepted psychological standards as defined by the American Psychological Association from time to time, and which has the least adverse impact on minority applicants. Consistent with this requirement defendants (a) shall examine all reasonably available alternative selection procedures on the subject of testing of police officer applicants, and (b) shall consult with industrial psychologists, psychometricians and/or others who have experience in the field of selection

court approval, and long-term hiring. The City objects most strenuously to the provisions concerning interim and long-term hiring, since these provisions involve the use of a quota.

---

testing, and preferably who have performed or have knowledge of analyses of the job of police officers.

7. Defendants may continue to use the current qualifications and selection criteria for police officer positions. However, no such qualification or selection criterion shall be a valid basis for or defense for failure to meet the interim hiring goals set out in paragraph 4, *supra,* unless the court has ruled that such qualification or selection criterion has been validated in accordance with the Uniform Guidelines and has determined that there is no further basis for continuing the interim goals.

8. Members of the plaintiff class shall be afforded the opportunity, at appropriate later proceedings, to show that they are entitled to an award of back pay and constructive seniority.

9. Plaintiffs are entitled to their court costs and reasonable attorneys' fees to date. The amount of such costs and fees shall be set by the court after a hearing. Costs and attorneys' fees for work done in the future shall be fixed in such manner as the court may determine.

10. Within thirty (30) days after the entry of this order, and every six (6) months thereafter, the defendant city shall submit to the plaintiffs the following reports:

(a) A list of its then current uniformed employees in the Police Department showing for each person: name, address, race or national origin, police station or other place of assignment, date of appointment, rank, and date such rank was achieved.

(b) The total number of uniformed personnel employed by the Police Department, by rank, race and natural origin.

(c) A list of all minority applicants for all vacancies, including date of application, names, addresses, and telephone numbers, whether the applicant was accepted or rejected and reason(s) for rejection.

(d) The name, address, and telephone number of any minority employee involuntarily terminated prior to the completion of the probationary period, and reason(s) for termination.

(e) List of all hires, promotions and voluntary and involuntary terminations showing race and national origin.

11. Defendants shall provide to plaintiffs, upon request, such other records or documents as are necessary to monitor compliance with this order.

12. The court retains jurisdiction of this action for such further relief or other orders as may be necessary or appropriate to enforce

In considering the District Court's order, we find it useful to distinguish between those aspects of the order that are designed to assure compliance with Title VII and those aspects that provide affirmative relief as a remedy for past discrimination. Compliance involves restricting the use of an invalid exam, specifying procedures and standards for a new valid selection procedure, and authorizing interim hiring that does not have a disparate racial impact.[25] Affirmative relief involves interim hiring at any ratio greater than what is necessary just to avoid a disparate racial impact and any required long-term hiring targets or ratios. Though standards in this difficult area are only beginning to emerge and have been a source of disagreement within this and other courts, we distill from the case law the following general principles, applicable to remedies for discrimination in entry-level hiring.[26]

1. As a general matter Title VII relief should at least assure compliance with the law. When it has been established that a selection procedure has been unlawfully used, an appropriate compliance remedy should forbid the use of that procedure, or its disparate racial impact, and may properly assure the establishment of a lawful new procedure. When it also appears that the employer has discriminated prior to the use of the challenged selection procedure, then it may also be appropriate to

---

and insure rights to equal employment opportunity within the New York City Police Department.

SO ORDERED.

25 An example would be an interim minority hiring ratio of 30%, where the minority percentage of the applicant pool or the relevant work force is also 30%. In this case, the District Court's provision for interim hiring specifies a minority ratio of a remedial nature, *i.e.*, greater than needed simply to avoid a disparate racial impact. We therefore consider this provision as part of affirmative relief.

26 The formulation of remedies for discrimination in promotion of employees requires even greater caution than is appropriate for entry-level hiring cases, since such remedies inevitably impact adversely identifiable employees who have committed themselves to a particular career track. *See Kirkland, supra,* 520 F.2d at 429; *Bridgeport Guardians, supra,* 482 F.2d at 1341.

fashion some form of affirmative relief, on an interim and long-term basis, to remedy past violations, see, *e.g., Prate v. Freedman,* 583 F.2d 42, 47 (2d Cir. 1978); *United States v. City of Chicago, supra,* 549 F.2d at 436-37; *Morrow v. Crisler,* 491 F.2d 1053 (5th Cir.) (*en banc*), *cert. denied,* 419 U.S. 895 (1974); *Bridgeport Guardians, supra,* 482 F.2d at 1340; *Carter v. Gallagher,* 452 F.2d 315, 331 (8th Cir. 1972) (*en banc*); *cf. Franks v. Bowman Transportation Co.,* 424 U.S. 747, 763-64 (1976) (Title VII authorizes broad remedial relief); *Albemarle, supra,* 422 U.S. at 418 (same). However, the form of such affirmative relief, especially the use of quotas, requires a most sensitive approach, see *Association Against Discrimination, supra,* 594 F.2d at 310-11; *Kirkland, supra,* 520 F.2d at 427-28; *Patterson v. Newspaper & Mail Deliverers' Union,* 514 F.2d 767, 775-76 (2d Cir. 1975) (Feinberg, J., concurring), *cert. denied,* 427 U.S. 911 (1976); *Bridgeport Guardians, supra,* 482 F.2d at 1340; *Vulcan Society, supra,* 490 F.2d at 398-99.

2. Initial consideration should be given to relief for the plaintiffs and those similarly situated, that is, Black and Hispanic applicants who took Exam No. 8155. While relief in Title VII cases need not necessarily be limited to the applicant class nor framed in specific relation to that class, their interests obviously deserve consideration. See *Castro v. Beecher,* 459 F.2d 725, 736-37 (1st Cir. 1972); *Carter v. Gallagher, supra,* 452 F.2d at 328-31.

3. Interim hiring provisions, for the period prior to use of a valid selection procedure, should be considered and formulated separately from long-term hiring provisions.

4. Since interim hiring provisions, where needed to satisfy immediate personnel requirements, are to be used prior to the development and approval of a valid selection procedure, such provisions cannot meet Title VII standards by demonstrated job relatedness. Therefore, one appropriate way to assure Title VII compliance on an interim basis is to avoid a disparate racial impact. This means selecting from among adequately qualified applicants either on a random basis, see, *e.g., Association Against Discrimination, supra,* 594 F.2d at 313, n.19, or according to some appropriately non-compensatory ratio

see, *e.g., Kirkland, supra,* 520 F.2d at 429-30; *Vulcan Society, supra,* 490 F.2d at 398-99, normally reflecting the minority ratio of the applicant pool or the relevant work force.

5. Any use of a hiring ratio during the interim period to compensate for prior discrimination, that is, a ratio greater than the minority percentage in the applicant pool or the relevant work force, should be imposed only upon clear evidence and appropriate findings of the need to redress demonstrated prior discrimination of long standing that has had a significant impact on minority employment. See *Association Against Discrimination, supra,* 594 F.2d at 312; *Patterson, supra,* 514 F.2d at 776 (Feinberg, J., concurring); *Vulcan Society, supra,* 490 F.2d at 398-99.

6. If a hiring ratio is imposed beyond the interim period in which a valid selection procedure is developed in order to reach a required long-term target, the justification for its use must be especially compelling.[27] See *Bridgeport Guardians, supra,* 482 F.2d at 1340. The prior discrimination warranting such a remedy must either be intentional, or it must plainly appear that significant discrimination has persisted for a substantial time. Gross disparity between minority employment and minority percentage in the relevant work force may imply such discrimination, especially when the minority employment is extremely low. Otherwise, the instances, impact, and duration of prior discrimination must be established.

With these principles in mind we turn to consideration of the order's provisions for compliance remedies and for affirmative relief.

---

27 This requirement of reaching a target should be contrasted with the far more modest use of a target figure merely to limit the extent of interim relief. The latter occurs when a remedy provides that the interim relief will continue until an acceptable selection procedure is developed *or* until a particular target figure for minority employment is achieved. In such a case, the target figure does not function as an absolute requirement; it simply serves as a means of assuring that the interim requirements will end at some point, even if development of a valid selection procedure is unduly delayed. The use of a target figure for this limiting purpose does not require the same compelling justification as a target figure prescribed as an absolute requirement, since it imposes no additional obligation on the defendant.

Compliance Remedies

Paragraph 2 of the order enjoins the use of Exam No. 8155 as a selection procedure, except in connection with the implementation of the interim and long-term hiring provisions. Deferring for the moment the exception concerning the permissible use of the exam, we readily affirm the District Court's prohibition against the unqualified use of the exam. The exam as used violated Title VII, and it is obviously appropriate to bar its continued use, except on an interim basis with adjustments that eliminate its disparate racial impact and thereby avoid its unlawful effect.

The order prescribes four requirements for the development and approval of a new selection procedure. We affirm the requirement, in paragraph 6 of the order, that the City make extensive efforts in its search for a new procedure, including consideration of "all reasonably available alternative selection procedures" and broad consultation with appropriate professionals.

We also affirm the procedural requirement in paragraph 4 of the order that the new selection device must be approved by the District Court prior to its use. Once an exam has been adjudicated to be in violation of Title VII, it is a reasonable remedy to require that any subsequent exam or other selection device receive court approval prior to use. See, *e.g., Bridgeport Guardians, supra,* 482 F.2d at 1339. This situation is to be contrasted with a case like *Guardians Ass'n v. Civil Service Commission,* 490 F.2d 400 (2d Cir. 1973) [*Guardians I* or "the '68-'70 exams case"], where the exams had not yet been found to be invalid. In that case the City was obliged only to show the new exam to plaintiffs and afford them an opportunity to criticize it.

However, we reject the District Court's principal substantive standard for approval of the new selection procedure to the extent that it requires any new procedure to be validated in accordance with the Guidelines and consistent with the APA Standards. As discussed in part III of this opinion, we have concluded that literal compliance with the Guidelines and with

professional testing criteria is not required by Title VII and can, in some instances, lead to results inconsistent with Title VII's explicit endorsement of "any professionally developed ability test." 42 U.S.C. § 2000e-2(h). We therefore conclude that the District Court, in determining the legality of a new selection procedure, should not require that it must conform in all respects to the Guidelines and the APA Standards; it will be sufficient if the new procedure conforms to the essential purposes of Title VII. We have endeavored to outline the extent to which the Guidelines are useful in carrying out those purposes and some of the respects in which excessive rigidity in application of the Guidelines may undermine those purposes. No all-encompassing formula is possible. The Guidelines remain useful as a source of guidance, but they need not be adhered to in every detail as if they were substantive regulations.

We also reject the District Court's substantive requirement, as expressed in paragraph 6 of the order, that the new selection procedure must have "the least adverse impact on minority applicants." This requirement appears to be an attempt to implement the principle expressed in *Albemarle* that once an employer has established the job-relatedness of a selection procedure that has a disparate racial impact, the plaintiff may still establish a Title VII violation by proving that "other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's interest in 'efficient and trustworthy workmanship.' " *Albemarle Paper Co. v. Moody, supra,* 422 U.S. at 425, quoting *McDonnell Douglas Corp. v. Green, supra,* 411 U.S. at 801. Of course, a decree may incorporate this principle into the standard for approving any new selection procedure, but the phrasing in paragraph 6 imposes a stricter and impermissible burden. To comply with Title VII a new selection procedure need not have the least adverse impact on minority applicants. That requirement would prohibit any exam with any disparate racial impact because random selection would always be a procedure with less adverse impact. What *Albemarle* contemplates, and what the decree may require, is that a selection procedure proposed

by the City may not be used if the plaintiffs can establish the existence of an alternative procedure with an equivalent degree of job relatedness and a lesser disparate racial impact.

Affirmative Relief

Our initial concern with the provisions of the order concerning affirmative relief arises from our uncertainty as to precisely what the District Court has required. The District Court's January 11 order states, and its January 23 revised opinion appears to require, that the City take the affirmative action of hiring 50% of entry-level police officers from among qualified Black and Hispanic applicants. But the opinion and the order seem to contain different provisions about the length of time for which this 50% minority hiring ratio is to apply. The opinion characterizes the affirmative action required as an "interim" measure. 484 F.Supp. at 799. In the context of Title VII testing cases, "interim" has meant the time period between the date of a decree and the subsequent use of a valid selection procedure. See, *e.g., EEOC v. Local 638, Sheet Metal Workers' Association,* 532 F.2d 821, 829 (2d Cir. 1976); *Kirkland, supra,* 520 F.2d at 423, 429-30; *Vulcan Society, supra,* 490 F.2d at 398. It is to be contrasted with a long-term or permanent hiring requirement, which specifies a minority composition of the employer's work force that must be achieved, even if a valid testing procedure has been developed and approved before the targeted minority composition is reached.[28]

In contrast to the District Court's opinion, its order, which is the operative document brought here for review, appears to

---

28 Though the District Court's opinion refers to affirmative action as an "interim" measure, it prescribes that it will last either until "discrimination has been totally eliminated" or until valid selection procedures are being used. (J. Supp. A. 27). The meaning of the first phrase is not entirely clear; however, even if it means a long-term objective of minority police employment equal to the minority percentage in the work force, the opinion still requires no more than an interim remedy because of the provision in the second phrase that the affirmative action can be terminated once a valid selection procedure has been approved.

require numerical quotas that continue beyond interim relief. Paragraph 3 states that the defendants "shall" seek to achieve minority (Black and Hispanic) representation in the Police Department "comparable to that of the minority composition of the labor force in the relevant hiring area," a representation stated to be at least 30%. Paragraph 4 also indicates that the required 50% quota hiring[29] may well last beyond the interim that ends with approval of a valid test. Paragraph 4 prescribes 50% minority hiring either until minority representation in the Department equals minority representation in the relevant labor force or until the District Court has both approved a valid selection procedure and, in addition, found that 50% quota hiring is no longer "appropriate." That such a finding might not be made until sometime after approval of a valid selection procedure and perhaps not until the long-term hiring goal has been reached is indicated by the specific reservation in Paragraph 4 of the plaintiffs' right to advocate the continued use of hiring quotas because "the continuing effects of past discrimination have not been eliminated."

In addition to creating uncertainty whether affirmative relief has been ordered on an interim or long-term basis, the record contains inadequate findings and, more significantly, inadequate evidence to support the hiring provisions of the order. The District Court determined that affirmative relief was warranted based upon conclusions concerning the prior employment practices of the defendants and their state of mind in preparing and using Exam No. 8155. The prior practices concern the defendants' use of an eligibility list compiled from the results of police exams given between 1968 and 1970. The continued use of the results of those exams after 1972, when Title VII was amended to include municipal employers, had previously been found to violate Title VII because the exams

---

29 The order unfortunately refers to the 50% quota as an "interim goal." It is not a goal at all. It is a procedure that the District Court has required to be used at least until the true interim goal—approval of a valid selection procedure—has been reached, and perhaps until the long-term goal—minority employment percentage equal to minority work force percentage—has been reached.

had a disparate racial impact and were not job-related. That conclusion had been reached in litigation concerning the Police Department's layoff policy, *Guardians Ass'n v. Civil Service Commission,* 431 F. Supp. 526 (S.D.N.Y.) [*Guardians II* or "the first layoff policy case"], *vacated and remanded for reconsideration,* 562 F.2d 38 (2d Cir. 1977), and reaffirmed in *Guardians Ass'n v. Civil Service Commission,* 466 F. Supp. 1273 (S.D.N.Y. 1979) [*Guardians III* or "the second layoff policy case"], *aff'd in part, remanded in part,* ____ F.2d ____ (2d Cir. July 25, 1980).[30]

The District Court grounded its decision to impose affirmative relief on the conclusion that the defendants designed Exam No. 8155 "either with a deliberate intention to discriminate against blacks and hispanics or with reckless disregard of whether the test would have that result." 484 F. Supp. at 798-99. This serious indictment of responsible city and police administrators is unsupported and indeed contradicted by the record. The conclusion is based virtually exclusively on the fact that the Police Department failed to assemble a valid eligibility list from the '68-'70 exams and failed again in 1979.[31] It would be contrary to Title VII's provision allowing the use of valid exams to hold that once an employer tries to construct such an exam and fails, any further failure to develop a valid exam

---

30 In 1972 the defendants gave still another exam whose validity has not been adjudicated. *See* 431 F. Supp. at 545 n.36. It is interesting to note, however, that even plaintiffs' well-known testing expert, Dr. Richard Barrett, acknowledged on cross-examination during the first layoff policy case that he could not be certain whether or not this exam is content valid. *Id.* n.37.

31 The only evidence referred to by the District Court, in addition to the invalidity of the '68-'70 exams, is the testimony of some police officers who warned the test makers that the test would disfavor minorities. (J. Supp. A. 26). Such a caution would be significant if the test makers had made no effort to satisfy Guideline standards. But when test makers have undertaken the elaborate process of job analysis and test construction revealed by this record, their willingness to disregard a prediction of disparate racial impact does not indicate that they lacked a good faith belief that their exam could nonetheless be shown to be adequately job-related.

constitutes intentional discrimination. Such a second attempt, by itself, is evidence only of a desire to make use of a technique that the law explicitly allows. Persistent use of exams with disparate racial effects would support an inference of intentional discrimination if proper test construction were not even attempted. But the record here indicates that the City's police and personnel officials made extensive efforts to understand and apply the Guidelines and develop a test they hoped would have the requisite validity.[32] Their failure entitles the plaintiffs to some relief, but does not justify a remedy based upon an unwarranted inference of deliberate discrimination.

In the absence of intentional discrimination, affirmative relief requires some demonstrated pattern of significant prior discrimination. There are no adequate findings concerning such a pattern, and the record lacks sufficient evidence on which such findings could be based. The District Court referred to the City's prior Title VII violation in using the eligibility list resulting from the '68-'70 exams and concluded that the minority "imbalance" on the City's police force is "directly caused by past and current discriminatory practices." 484 F. Supp. at 799. Obviously Exam No. 8155 has had no significant effect upon the current minority proportion of the police force, because its results have been used to select only one class for the training academy. As to the use of the eligibility list from the '68-'70 exams, there are no findings and no evidence to indicate the extent to which use of that list has affected the minority proportion of the police force. The first layoff policy case provides some data as to the numbers of Whites and minority members hired as a result of two of the '68-'70 exams, 431 F. Supp. at 552-53, Tables 4 and 6, but even

---

32 It is true that the City's previous unsuccessful attempt to design an exam on its own renders somewhat questionable its apparent enthusiasm for another "in-house" effort. However, the decision to produce Exam No. 8155 "in-house" may well have been motivated by a bureaucratic preference for internal procedures, a need to save money, a naive self-confidence, or simply a desire to try again. None of these motives is a basis for inferring a conscious intention or even a reckless willingness to violate the law.

with that data, the record in this case does not disclose the minority percentage in the police department before and after the '68-'70 exams. Nor is there any evidence of the impact of hiring resulting from Exam No. 3014, administered in 1973, the validity of which has not been challenged. The '68-'70 exams undoubtedly made some contribution to the current racial imbalance of the police force, but the record does not contain even estimates of how the hiring prior to the 1973 exam currently affects the composition of the police force. Plaintiffs have failed to prove that prior use of discriminatory exams has created a situation warranting affirmative relief.

In the absence of proof of the specific impact of such prior discrimination, the only probative evidence in the record is that the current minority proportion of the police force is 12.7% compared to a relevant work force percentage of at least 30%. That is cause for some concern, but does not reveal the flagrant disparity shown in prior cases where long-term hiring quotas were in issue. *Cf. Association Against Discrimination, supra,* 594 F.2d at 308 (minorities constituted 0.2% of employees, 41% of population; quota vacated for reconsideration and findings); *EEOC v. Local 14, International Union of Operating Engineers,* 553 F.2d 251, 256 (2d Cir. 1977) (minorities constituted 2.8% of union members, at least 16.2% of relevant labor force; judgment including quota vacated for further findings); *Patterson, supra,* 514 F.2d at 770, 772 (minorities constituted 2.45% of union and union-affected job-seekers, 30% of relevant labor force; quota sustained); *Bridgeport Guardians, supra,* 482 F.2d at 1335 (minorities constituted 3.6% of employees, 25% of population; hiring quota sustained). If the disparity between existing minority employment and relevant work force percentage were extreme and long-standing, that circumstance alone might justify some affirmative relief, especially if minority employment were low. But where, as here, the disparity is not extreme and minority employment is not insubstantial, an affirmative hiring remedy must be based on detailed findings, supported by evidence, that there exists a pattern of prior discrimination warranting such relief. *Cf. Association Against Discrimination, supra,* 594 F.2d at 312-13; *Kirkland, supra,* 520 F.2d at 427-28.

We therefore conclude that the affirmative hiring provisions of the order must be set aside. This will require elimination of the affirmative hiring quota of 50%, both as interim and long-term relief, and elimination of the long-term hiring goal of 30%, a goal that obviously could be achieved only by affirmative hiring at a ratio above the minority percentage of the relevant work force. The only hiring remedy justified by this record is a compliance remedy, one designed to make sure that the City complies with the requirements of Title VII in making appointments to the police force. Such a remedy should permit the City, in the interim period prior to development of a new, valid selection procedure, to use the results of Exam No. 8155 in a way that avoids any disparate racial impact. This means selecting candidates from the eligibility list subject to the minority proportion of either the applicant pool or the relevant work force, a determination to be made upon remand.

To accomplish such interim hiring the City may assemble a minority pool and a majority pool of qualified candidates from the eligibility list. In assembling these pools, the City may use a cut-off score somewhat lower than 94, a score that was originally determined by the City's manpower needs, rather than by an independent estimate of adequate ability. Within the majority and minority pools, the City may choose candidates, maintaining the requisite proportion of minority candidates and taking into account those already hired as a result of this exam. The City is not obliged to hire on an interim basis, but it should have the option of doing so in order to meet its manpower needs.

We remand this case to the District Court for the entry of a revised decree consistent with this opinion. Pending the entry of that decree, we continue in effect the provisions of the stay order we previously entered, under which the City is afforded the option of hiring from those who scored 94 or above on Exam No. 8155 provided such hiring achieves a minority ratio of 33%, taking into account those already hired as a result of this exam.

Affirmed in part, vacated in part, and remanded.

SIFTON, D.J., *concurring:*

I concur in the result and in the reasoning of Judge Newman's thorough opinion with the exception of his conclusion that Exam No. 8155 adequately tested a representative sample of the skills, knowledges and abilities required for police work in New York City. While I join in rejecting the argument of the United States as *amicus,* that the requirement of representativeness means that *all* the knowledges, skills and abilities needed for police work must be tested, each in its proper proportions, I disagree with the conclusion that an exam which contains omissions as extensive as the present exam meets the "representativeness requirements to an adequate degree."

As Judge Newman correctly points out, Exam No. 8155 failed to test for human relations skills despite the City's job analysis which found, not surprisingly, that such skills constitute a significant part of police work in New York City—sufficiently important to warrant devoting 30 of the exam's 100 questions to determining whether the applicants possessed them and in what degree. Unfortunately, as Judge Newman so well explains, the portion of the exam which attempted to test for human relations skills in fact simply replicated other portions of the exam which tested the applicants' abilities to remember details, fill out forms, and apply general principles in specifically described factual settings. In my view, an exam which does not test for a skill or ability determined to constitute close to one-third of the talents required for success on the job cannot be called representative in any meaningful sense.

The legal effect of excusing a lack of representativeness of these dimensions, as Judge Newman's opinion appears to do, because the record does not contain evidence that a test for the relevant skills or abilities is "readily available and realistically feasible," appears to me, with all due respect, to shift to plaintiffs the burden of explaining the exam's racially disparate impact. The practical effect of validating a test for New York City police work which does not examine for human relations skills is to leave out of the entry level employment decision an area of qualifications in which minority groups would, one must assume, perform well despite educational and other

deprivations. The immediate consequence of the decision is that high performance in the area of human relations skills will not be available to alter the overall assessment of applicants who perform less well in other areas.

Nor will refinements in rank-ordering or in the determination of an appropriate cut-off score overcome the effect of an omission of these dimensions. Selection at random or in rank order without the benefit of any assessment of the applicants' performance in one large area of police work will reflect the same racial disparities as exist in the pool from which random selection is made or in the ranks established by the unrepresentative test.[1]

---

1 Of course an unrepresentative exam—as is the case with exams at other stages of the qualification process for New York City police work—can be administered on a pass/fail basis as part of a larger qualification process provided the cut-off score represents that level of skills shown to be the level below which an applicant is disqualified for police work. There is, in other words, no objection on grounds of discriminatory impact to a test which eliminates those whose skills and abilities in the areas actually tested for are so low as to disqualify the applicant no matter how well he or she might perform in other untested areas.

Appendix B
Order of the Second Circuit

UNITED STATES COURT OF APPEALS

SECOND CIRCUIT

At a Stated Term of the United States Court of Appeals, in and for the Second Circuit, held at the United States Court House, in the City of New York, on the fifteenth day of January, one thousand nine hundred and eighty.

Present:

HON. WALTER R. MANSFIELD
HON. JON O. NEWMAN

Circuit Judges

HON. CHARLES P. SIFTON

District Judge

80-7027

THE GUARDIANS ASSOCIATION OF THE NEW YORK CITY POLICE DEPARTMENT, INC., THE HISPANIC SOCIETY OF THE NEW YORK CITY POLICE DEPARTMENT, INC., NYDIA I. DIAZ, JAMES MICHAEL HIDALGO, WILFRED CEBELLERO, ANDRE LOPEZ, REINALDO SALGADO, DENISE SANTO, DEBORAH HOLMES and PAMELA OBEY, individually and on behalf of all those similarly situated,

Plaintiffs-Appellees,

—v.—

PRECEDING PAGE WAS BLANK

CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK, DEPARTMENT OF PERSONNEL OF THE CITY OF NEW YORK and THE NEW YORK CITY POLICE DEPARTMENT,

Defendants-Appellants.

A petition for a rehearing having been filed herein by counsel for the plaintiffs-appellees.

Upon consideration thereof, it is

Ordered that said petition be and hereby is denied.

s/ A. DANIEL FUSARO
A. Daniel Fusaro, Clerk

United States Court of Appeals
Second Circuit
Filed: Jan. 15, 1981
A. Daniel Fusaro, Clerk

Appendix C
Opinion of the District Court

UNITED STATES DISTRICT COURT

SOUTHERN DISTRICT OF NEW YORK

79 Civ. 5314 (RLC)

THE GUARDIANS ASSOCIATION OF THE NEW YORK CITY POLICE DEPARTMENT, THE HISPANIC SOCIETY OF THE NEW YORK CITY POLICE DEPARTMENT, NYDIA I. DIAZ, JAMES MICHAEL HIDALGO, WILFRED CEBOLLERO, ANDRE LOPEZ, REINALDO SALGADO, DENISE SANTOS, DEBORAH HOLMES, and PAMELA OBEY, individually and on behalf of all those similarly situated,

Plaintiffs,

—against—

CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK, DEPARTMENT OF PERSONNEL OF THE CITY OF NEW YORK, and THE NEW YORK CITY POLICE DEPARTMENT,

Defendants.

APPEARANCES:

THE PUERTO RICAN LEGAL DEFENSE AND EDUCATION FUND, INC., 95 Madison Avenue, New York, New York 10016, Attorneys for Plaintiffs.

By: KENNETH KIMERLING, ESQ. and PETER BIENSTOCK, ESQ., of Counsel.

ROBERT B. FISKE, JR., United States Attorney for the Southern District of New York, One St. Andrew's Plaza, New York, New York 10007, Attorney for *Amicus Curiae* United States of America.

By: NANCY E. FRIEDMAN, RICHARD N. PAPPER, and DENNISON YOUNG, JR., Assistant United States Attorneys of Counsel.

IRA H. LEIBOWITZ, ESQ., 585 Stewart Avenue, Garden City, New York 11530, Attorney for *Amicus Curiae* Policewomen's Endowment Association.

ALLEN G. SCHWARTZ, Corporation Counsel of the City of New York, Municipal Building, New York, New York 10007, Attorney for Defendants.

By: JUDITH A. LEVITT, ESQ., MAUREEN MCCABE, ESQ., and STEVEN GOLDBERG, ESQ., of Counsel.

CARTER, *District Judge:*

REVISED OPINION[1]

Summary of Proceedings

On June 30, 1979, Examination No. 8155 was given to provide an eligibility list of applicants for appointment as police officers on the New York City police force. Approximately 36,797 applicants took the examination. The police department had indicated that it could absorb an eligibility list

1 This opinion replaces the opinion filed on January 11, 1980, and that opinion is withdrawn.

of some 12,000 candidates over the next four years. Because the 12,000th candidate scored 94, that became the cut off point for placement on the eligibility list. Roughly 13,000 persons qualified with scores of 94 or above. The city proposes to select applicants for the police force by their ranked order of combined examination scores and veterans' preference points until the list is exhausted. In November, 1979, defendants hired 415 police recruits from the eligible list: 318 white males, 49 white females, 11 black males, 3 black females, 29 hispanic males, 2 hispanic females and 3 oriental males.

Plaintiffs, the Guardians Association of the New York City Police Department, Inc., the Hispanic Society of the New York City Police Department, Inc. and Nydia I. Diaz, James Michael Hidalgo, Wilfred Cebollero, Andre Lopez, Reinaldo Salgado, Denise Santos, Deborah Holmes and Pamela Obey, individual black and hispanic applicants who either received a passing grade too low on the eligibility list to afford an expectation of appointment in the forseeable future, or who scored below the qualifying 94, instituted this litigation. They challenge the legality of the examination and seek a preliminary and permanent injunction barring the planned use of the test's results. Defendants, the Civil Service Commission, the Department of Personnel and the Police Department of the City of New York, are responsible for the preparation, administration and use of the examination.

As indicated, the eligibility list has already been utilized, and defendants propose to hire another class of recruits from the list on Monday, January 14, 1980. With the parties' consent the hearing on the preliminary injunction was consolidated with a trial on the merits which took place on November 13, 14 and 15.

At the hearing defendants presented testimony indicating how the examination had been devised and structured. Plaintiffs presented similar testimony by individual participants in the preparation of the test, as well as evidence concerning the adverse effect on minorities of the examination and of use of an eligibility list with ranking based on the candidates' scores. Both sides presented expert testimony concerning the test's

validity. The parties filed their briefs in support of their contentions on or about December 3, 1979, but neither side filed proposed findings of fact and conclusions of law until on or about December 13, 1979. On December 10, 1979, the government filed a brief *amicus curiae* asserting that Examination No. 8155 was unlawful, and thereafter participated as *amicus curiae* in the proceedings.

On December 17, 1979, the court advised the parties orally that it was of the view that the examination and the rank order selection of designated eligible candidates violated Title VII of the Civil Rights Act, 42 U.S.C. § 2000e *et seq.*, and that no appointments from the list in ranked order should be made. The court anticipated that this oral indication of its determination would enable the city to make temporary adjustments in the selection process to bring the contemplated January appointments into compliance with the law by taking affirmative action to eliminate discrimination against blacks and hispanics. At a subsequent hearing on defendants' order to show cause and motion for reconsideration, counsel for defendants stated that defendants would not modify the selection process by the appointment of black and hispanic applicants out of rank order, and that they were determined to utilize the discriminatory listing. The matter was then heard by the Court of Appeals on defendants' petition for a writ of mandamus. The Court of Appeals ordered this court not to bar the city from making the contemplated appointments without findings of fact and conclusions of law being filed in compliance with Rules 52(a) and 65(d), F.R.Civ.P. The court was further ordered to refrain from enjoining defendants from selecting police recruits from the existing list until 48 hours after such findings and conclusions had been filed.

On January 11, 1980, a hearing on relief was held. The Policewomen's Endowment Association was allowed to intervene as *amicus curiae*. Plaintiffs offered testimony on the black-hispanic make-up of the relevant labor pool from which applicants for employment with the New York City police force were expected to come. Defendants have qualified for immediate appointment on January 14, 1980, some 380 candidates, of

whom 280 are white males and 62 are white females, 9 are black males and 5 are black females, 20 are hispanic males and 3 are hispanic females, and 1 is an oriental male.

Preparation of Examination No. 8155

The development of the test and the job analysis employed was described at the November hearings by Esther Juni, administrative staff analyst of the department of personnel and chief of the police unit of the criminal justice task force. Juni disclaimed expertise in both testing and police work. This was only the second entry level examination she had ever worked on, the first being an examination for police administrative aides manning the 911 system where the skills tested were basically clerical and communicative. Moreover, no outside experts in testing had been brought in to aid defendants in devising a test in compliance with the Uniform Guidelines on Employee Selection Procedures ("Uniform Guidelines") promulgated by the Equal Employment Opportunity Commission, Department of Justice, Department of Labor and the United States Civil Service Commission, 5 C.F.R. § 300.103(e), 28 C.F.R. § 50.14, and 29 C.F.R. § 1607, effective September 25, 1978.

Juni's testimony was hardly a model of clarity. It was often confusing, contradictory and obscure. However, from her testimony and the report outlining the procedures followed in job analysis and in preparation of Examination No. 8155, which I believe she testified she prepared or whose preparation she supervised, the following facts were gleaned.

Preparation of the test began in the fall of 1978 when, as a first step, the department of personnel asked the police department to supply 10 staff analysts to interview 49 police officiers and 49 of their supervisors. These officers, in Juni's words, were chosen on the basis of "good job performance record, their ethnicity, the sexual composition (sic), types of duty they performed, the tours they worked on, the boroughs they worked in and it was made to be a comprehensive analysis of the positions of the police officers" (Tr. 9-10). The officers

were asked to identify the tasks performed by police officers, their frequency and importance, the amount of time spent thereon, and the knowledge, skills and abilities (labeled KSA's)[2] necessary to perform each identified function. By this process some 71 tasks were identified.

Presumably this group of officers was chosen as "job knowledge experts,"[3] but except that they were apparently selected to comprise a cross section of good performers, we do not know their level of expertise or indeed whether they qualify as job knowledge experts[4] as that term is used in the department of personnel manual on job analysis (*"Procedures"*). There is no evidence that Juni or anyone else examined the officer-consultants on this question before asking them to define the tasks police perform.

Next, a panel of police officers was chosen to refine the identified tasks further. Again, except that the panel was chosen by officers on the police force, using the same criteria Juni described above, we were not advised whether the panel members possessed any special job knowledge expertise warranting their selection over others or even insuring their professional competence to do the task assigned. At any rate, the

---

2 While Juni's testimony implied that the KSA's the examination was structured to measure were knowledge, skills and abilities, Dr. Finkelman testified that the KSA's did not measure knowledge but only abilities and skills needed to perform the police officer's function.

3 This is a term used in the manual entitled "Department of Personnel Procedures for Job Analysis and Preparation of Notice of Examination" (*"Procedures"*). On page one, the manual describes itself as a "guide . . . to provide examiners of the Department of Personnel with a tool for developing job-related and content-valid examinations for appointments and promotions. . . ." *Procedures* goes on to state that the instructions embodied described in the manual are designed to be consistent with the Uniform Guidelines, American Psychological Association's (APA) Standards and recent court decisions. *Ibid.* Juni testified that *Procedures* was in effect at the time the instant test was being structured.

4 The report indicates the length of service of the various police officers. Long experience, however, does not necessarily translate into job knowledge expertise.

panel reduced the 71 identified tasks to 42. Thereafter a questionnaire was sent to 5,600 officers for their rating of the 42 tasks in terms of frequency, time spent and level of importance. Some 2,500 responses were received. At the same time, observers from John Jay College, watching police officers at work, recorded their observations of 28 of the identified 42 tasks. These observations were then correlated with the officers' responses to the questionnaires as to the frequency with which tasks were performed. The questionnaires were fed into a computer programmed to take data on time, frequency and importance, average them and produce a list of the 42 tasks in hierarchical order based on these three factors.

The tasks were next divided into 5 clusters of tasks "that seem to us to go together on a rational basis," "that looked somewhat similar," "had some kind of common factor" (Tr. 19). A panel of officers knowledgeable about the tasks involved was chosen for each cluster. If, for example, one of the clusters was the arrest process, police officers very familiar with the arrest process, "meaning they were in a precinct that maybe made a large number of arrests, etc." were chosen (Tr. 20). These panels were to determine the KSA's needed for each task within the cluster. Again the job knowledge expertise of these people is not identified or defined, nor was that factor determined by Juni or anyone else from the department of personnel. Juni's report refers to this group as "job knowledge experts" without further elaboration. The KSA's to be assigned to each task by the panels were characterized as filling out forms; human relations, including communication techniques; understanding written instructions and applying appropriate procedures; comprehending and applying sections of the law; and recalling details and facts. The KSA's were never more specifically defined, although the human relations KSA's were described as interacting with people.

The 5 panels never exchanged ideas or concepts, and we do not know whether the description and identification of the KSA's were the same or different from panel to panel. The KSA's were then rated by the 5 panels as to the relative importance each had in the successful completion of each task

within the assigned cluster. John Galea, one of the officers detailed to one of the panels assigning the relevant KSA's to each task, testified, however, that his group made no effort to differentiate the human relations KSA components assigned to different tasks. The panels' conclusions were fed into a computer which multiplied the weight of each task by the weight of the KSA's necessary to accomplish it. Based on the weighted result, the test plan was devised with an appropriate percentage of the examination being allocated to each of the 5 KSA's.

Then police officers were asked to propose questions covering the 5 KSA's. According to Juni the police department was asked to choose for this assignment "police officers who were familiar with the job of police officer" (Tr. 24). Some 10-11 officers were given this responsibility. They were not shown the KSA lists or the tasks with KSA weight accorded to them. Gerald Burkhalter, one of this group, testified that he was asked by an officer in the personnel bureau of the police department to participate. He was told they were looking for people of various ethnic groups and that he had patrol experience and a college background. Presumably these factors and his being black qualified him for the assignment. Between December, 1978 and June, 1979, he spent each Friday writing questions. The first two months he wrote questions based on materials from the Police Academy. He was not shown the tasks until he was over two months into the assignment and never saw the KSA's. He stated that he raised the issue of the adverse impact of written tests on minorities several times with Juni and in the equal opportunity office of the police department where he worked (Tr. 408). Although police officers proposed some of the questions, a good proportion of the questions were written by employees of the department of personnel.

A panel of police officers reviewed the questions to determine whether the police procedures described in the examination were consonant with current practices. Nicholas Estavillo, asked to work with this panel, joined it at its second meeting. The group was told what the questions dealt with, and discussed whether there was any conflict between the questions

and police department procedures. Estavillo concluded that what was being tested was reading comprehension, and expressed this view to Juni and others on several occasions (Tr. 440). Department of personnel employees reviewed all of the proposed questions to make certain they were stated clearly, simply, and gramatically, but these employees had no clear idea of what KSA's the questions they clarified and simplified purported to test. There was no final review by police or department of personnel officials to determine whether the questions as finalized actually tested for the KSA's they purported to measure.

The examination was divided into 100 multiple choice questions. Questions 1-15 dealt with recalling details and facts; questions 16-24 with filling out forms; questions 25-38 with comprehending and applying sections of the law; questions 39-62, 64-68, and 71-74 with understanding written instructions and applying appropriate procedures; and all the rest with human relations, including communications techniques. It was decided not to pre-select a passing score, but to qualify the 12,000 top scores.

Juni testified that she did not measure police work behavior but, rather, the KSA's necessary to perform a police officer's functions; that the test did not mirror the tasks of officers on the job, but the KSA's needed for the job (Tr. 48). She conceded that the work product that is the police officer's basic function was not completely described, and testified that while some of the tasks concern work product, work product must be inferred from others (Tr. 50). When asked for an operational definition of the KSA's, she replied that the KSA's are themselves simply stated. The KSA's refer to one or many tasks, and "the tasks themselves are operational definitions in the sense that the tasks themselves are things" (Tr. 51).

At one point we were told that none of the 42 tasks relate to training at the academy (Tr. 67); subsequently we were advised that the police recruits are trained in the tasks at the academy, and that the 42 tasks had to be used as a basis for isolating the KSA's (Tr. 68). What was being tested, Juni said, was whether the candidate could apply the information given him. Proce-

dures were taught at the academy, but not how to apply them. She conceded, however, that if "you understand procedures, you know the way it is applied" (Tr. 70).

After stating what the 5 KSA's were said to be, she could describe them no further. Police officers were the ones who knew that. They also were the ones who could tell us whether the level of difficulty and complexity or abilities needed differ among tasks. She could not tell whether the examination questions actually measured the relevant KSA's. If the police approved a human relations question, that meant they believed the question measured the human relations KSA's, but she could not independently verify this.

The KSA's were identified by one group, questions were written by a second group, and evaluation was done by a third group. No panel was shown the basis for another panel's KSA determinations, because it was felt that the stated KSA's would be understandable to police officers without further elaboration. The KSA's were said to differentiate the good from the bad officer, but how that was done was not explained.

The next witness, Dr. Jay Finkelman, defendants' expert, was impressive, but his central thesis was unpersuasive. Dr. Finkelman argued that the test was content valid, and that one could validate a test for content if it tested for work behavior or for the KSA's necessary for work behavior. He defined the KSA's as present capacity to perform a certain function and skills requisite to work behavior (Tr. 138), but he could not say whether the various panels involved had been operating pursuant to his definition (Tr. 141). While he did not completely understand what a particular task entailed, or what the police officer meant in listing the KSA's, he asserted that he, as an expert, could testify as to the test's content validity (141), as long as he understood the procedures through which the KSA's were derived, and the manner in which the KSA's determined the final test (Tr. 209).

He agreed that a content validation model was inappropriate for testing constructs (Tr. 149), but argued that the Uniform Guidelines were not violated by the test's overlapping the training the successful candidate would receive prior to em-

ployment. Moreover, he could not see how what was being tested coul be separated from training, since the tasks comprising the test were common tasks performed by the police on the job. The skills measured—for example the human relations KSA's—were not the human relations skills themselves, but a procedure ability to follow guidelines and instructions (Tr. 197); thus, he concluded, the examination was really a paper and pencil test of procedure-following ability.

Prudence Opperman testified that she had analyzed the test to determine its level of reading difficulty. Based on two formulas (Dale-Chall and Fry), she concluded that the level of reading difficulty was no higher than the 12th grade (Tr. 270). On cross examination, she conceded that some of the questions did require college level reading skills (Tr. 286-289), but stated that people reading at the 12th grade level can handle some material at the college level. She characterized this as "loading" (283).

Plaintiffs' two experts, Dr. Richard Barrett and Dr. James J. Fitzpatrick, both of whom helped draft the Uniform Guidelines, testified that the examination was akin to an aptitude test and that it had not been shown to be content valid. They testified that it did not measure job skills, but the skills of following given instructions and applying them to a hypothetical situation. The test measured constructs, not work behavior; further, it was the type of test that typically disfavors minorities. Measured against the standard of the American Psychological Association ("APA") and the Uniform Guidelines, the test was said to be deficient.

They both contended that it had not been established that the examination measured abilities which differentiate a good from a poor police officer. For that reason, they felt rank order selection was inappropriate. While conceding the difficulty of constructing a test for police officers that could be validated as having job relatedness, they both testified that if tests such as this were used, the adverse impact on minorities should be eased by maintaining separate lists for blacks, hispanics and whites, and selecting from these separate lists in a way that would insure equal opportunity to the minorities. They were the more credible witnesses.

The Examination Results

The test as given presented a series of issues or problems that police purportedly met on the job and for which the applicants were asked to select the correct responses. Of the 36,797 persons taking the test, 16.7% (6,142) were black, and 14.2% (5,239) were hispanic. 37% (13,749) of all applicants qualified with scores of 94 or above. Of this group only 7.6% (1,046) were black and 7.8% (1,074) hispanic. The list is roughly 84% white. Thus, although blacks and hispanics comprised 30.9% of those taking the test, they constituted only 15.4% of those on the eligibility list. To put it another way, 46% of the white applicants (9,094 of 19,798) made the eligibility list with scores of 94 or above, while only 17% of the black applicants (1,046 out of 6,142) and 20.5% of the hispanic applicants (1,074 out of 5,239) were similarly successful. Plaintiff's Ex. 3, applying standard deviation analysis, reveals that the disparity between the number of blacks and hispanics who could be expected to qualify on Examination No. 8155 and the number who actually qualified is some 39 standard deviations.

Hearing on Relief

On Friday, January 11, a further hearing was held on relief. Plaintiff presented, through Dr. Mark Killingsworth, a labor economist, evidence based on *Survey of Income and Education,* 1976, United States Census Department ("SIE"), to the effect that the relevant labor pool of those who by virtue of age, residence and education would be expected to be applicants for the job of police officer was 34.9% black and hispanic. The geographic area used by Dr. Killingsworth was New York City, Rockland, Westchester, Nassau and Suffolk Counties. The group included persons aged 17-29 with 12 or more years of education. He utilized the same geographic base as the area from which those who took Examination No. 8155 came, except that he excluded Orange and Putnam Counties. However, only .31% of the applicants who took the test at issue here came from Orange and only .10% came from Putnam. Defendants object to the pool as defined. They presented no

evidence to the contrary, but requested permission to file whatever documents they desired to present after the hearing. Defendants have filed an affidavit of Martin Oling.[5]

Determination

I

Section 4D of the Uniform Guidelines states that any "selection rate for any race, sex or ethnic group which is less than four-fifths ($^4/_5$) (or eighty percent) of the rate of the group with the highest rate will generally be regarded by the Federal

---

5 Defendants were given permission to make further submissions after the January 11, 1980 hearing on relief. Defendants have now submitted an affidavit, dated January 15, 1980, from Martin Oling, a statistician employed by the New York City department of planning, challenging Dr. Killingsworth's analysis. Accompanying the affidavit are various exhibits and charts from the 1970 United States census, extrapolations apparently prepared by the affiant from that data, extracts from Technical Documentation, 1976 Survey of Income and Education ("SIE"), Computer Microdata File, with appendices dealing with standard error computations and various other documents related to 1976 SIE.

I cannot consider this affidavit and the exhibits attached to it. The parties had been informed on December 17, 1979, that the court desired to hear live testimony on the issue of relief. Plaintiffs advised the court that defendants knew by December 27, 1979, that they planned to call Killingsworth and were given a computer print-out from which the witness was expected to testify. At the January 11 hearing, defendants' counsel cross examined Killingsworth at some length. Plaintiffs allege that Oling was in the courtroom at the time. The court cannot attest to that, but there were a number of people present who appeared to be assisting defendants' counsel, and she interrupted her cross examination of Killingsworth a number of times to consult with several such persons.

Acceptance of Oling's affidavit with its voluminous exhibits is not appropriate without first allowing plaintiffs the opportunity to cross examine him, as defendants were permitted to cross examine plaintiffs' expert. Accordingly, defendants' affidavit is not properly before the court and cannot be considered at this time. This does not bar the defendants, however, at some future time, from requesting a hearing to present evidence establishing that the court's order should be modified on the ground that blacks and hispanics do not constitute 30% of the relevant labor pool from which the police department is expected to draw its applicants.

enforcement agencies as evidence of adverse impact. . . ." *See, e.g., United States v. City of San Diego,* 20 E.P.D. ¶20,154 (S.D. Cal., 1979); *United States v. City of Montgomery,* 19 E.P.D. ¶9239 (M.D. Ala. 1979). The Uniform Guidelines are, of course, entitled to deference and should be followed, unless the employer demonstrates some cogent reason to the contrary. *Cf. Griggs v. Duke Power Co.,* 401 U.S. 424, 433-34 (1971); *Albemarle Paper Co. v. Moody,* 422 U.S. 405, 431 (1975). *United States v. City of Chicago,* 549 F.2d 415, 430 (7th Cir.), *on remand,* 437 F. Supp. 256 (N.D. Ill.), *aff'd,* 567 F.2d 730 (7th Cir. 1977), *cert. denied,* 436 U.S. 932 (1978), 434 U.S. 875 (1977). Moreover, the Supreme Court in *Hazelwood School District v. United States,* 433 U.S. 299, 308 n.17 (1977), and *Castaneda v. Partida,* 430 U.S. 482, 496 n.17 (1977), has recognized that "if the differences between the expected value and the observed number is greater than two or three standard deviations then the hypothesis that the [selection] was random would be suspect to a social scientist." *Ibid.* Here, the uncontroverted evidence is that disparity between the number of blacks and hispanics expected to qualify on Examination No. 8155 as opposed to the number who qualified in fact is some 39 standard deviations. As the Court stated in *Castaneda* such a difference is statistically significant and could not have occurred by chance. This result compounds and perpetuates a pattern of disparity in minority group representation on the New York City police force. See Def. Ex. A-1 (Report of Test Development), Tables 5 and 6 at 15, which reveal that as of June 30, 1978, of some 20,926 members of the police force, 1,860 (8.9%) were black, and 799 (3.8%) were hispanic. Census figures for 1978 show that the relevant labor force of the New York Standard Metropolitan Statistical Area (SMSA) is more than 30% black and hispanic. See State, County and Selected City Employment and Unemployment Figures, January-December, 1978, Bureau of Labor Statistics.

Where it has been shown, as here, that the selection process qualifies applicants for hire or promotion in a racial pattern significantly different from that of the pool of applicants, *see McDonnell Douglas Corp. v. Green,* 411 U.S. 792, 802 (1973),

a prima facie case of descrimination has been established. *Albemarle Paper Co. v. Moody, supra,* 422 U.S. at 425. A prima facie case may be established by evidence of statistical disparities alone. *Dothard v. Rawlinson,* 433 U.S. 321, 329 (1977). The burden then shifts to the employer to demonstrate that the process used is job related or, to put it more precisely, has a "manifest relationship to the employment in question." *Griggs v. Duke Power Co., supra,* 401 U.S. at 432; *Guardians Association v. Civil Service Commission of New York,* 431 F. Supp. 526, 538 (S.D.N.Y.) (Carter, J.), *vacated and remanded on other grounds,* 562 F.2d 38 (2d Cir. 1977) (*"Guardians* II"), *on remand,* 466 F. Supp. 1273 (S.D.N.Y. 1979), (Carter, J.) *app. pending,* No. 79-7377 (2d Cir.) (*"Guardians* III"); *Vulcan Society of the New York Fire Department v. Civil Service Commission,* 360 F. Supp. 1265, 1268 (S.D.N.Y.) (Weinfield, J.), *aff'd in part, remanded in part on other grounds,* 490 F.2d 387 (2d Cir. 1973). If the employer meets this burden, the plaintiff is then entitled to show that alternative processes would meet the employer's needs without the adverse racial impact that the complained of process inflicts on the minority group. *Albemarle Paper Co. v. Moody, supra,* 422 U.S. at 425.

II

We now examine the record to determine whether the test processes employed meet the required standard of a "manifest relationship" to the job of police officer in the City of New York. Since 1972, when Title VII first became applicable to municipalities, the police department of New York City has been under attack in courts for selection procedures alleged to be discriminatory against blacks and hispanics, *see Guardians Association of New York City Police Department, Inc. v. Civil Service Commission,* 72 Civ. 928 (S.D.N.Y.) (Ryan, J.), *aff'd* 490 F.2d 400 (2d Cir. 1973) (*"Guardians* I"); *Guardians* II, *supra; Guardians* III, *supra;* and against women, *see Acha v. Beame,* 401 F. Supp. 816 (S.D.N.Y. 1975), *rev'd,* 531 F.2d 648 (2d Cir. 1976), *on remand,* 438 F. Supp. 70 (S.D.N.Y. 1977) (Duffy, J.), *aff'd,* 570 F.2d 57 (2d Cir. 1978). Examination No. 8155 is the first examination for entry level police officer given

since 1973, and the above cases surely put city authorities on notice that this examination would be tested under Title VII strictures.

One of the accepted indicia of a lack of job relatedness is proof of a flawed job analysis. In *Vulcan Society v. Civil Service Commission, supra,* 490 F.2d at 396, this circuit indicated that the poorer the preparation of the test, the greater the need to demonstrate its "manifest relationship" to the job.

In *Guardians* II, the job analysis for examinations given prior to 1972 had come under heavy attack. Preparation for Examination No. 8155 began in the fall of 1978. At about that time, the department of personnel published *Procedures,* a guide setting out the requisites of job analysis for an examination which would conform to APA's professional standards and satisfy the Uniform Guidelines. Esther Juni was given the responsibility for designing the 1979 test. She took some of the steps required by *Procedures,* APA standards and the Uniform Guidelines but deviated therefrom in some respects. Unfortunately, her departures fatally flawed the test's preparation and in essence reduced that effort from the required professionalism to the level of a homemade test.

The most fatal flaw of all is that Juni as the test designer lost control of various critical aspects of the test's preparation. She went to police officers for help in defining the work behaviors and abilities in a policeman's job performance. While *Procedures* approves of this practice, it is made clear that the test maker's responsibility is to verify for herself that the KSA's or behaviors to be tested are in fact those actually performed on the job. The police were asked to furnish officers to describe the tasks performed, to define the abilities needed and to rate them in terms of frequency, importance, and time spent, and to frame appropriate questions. She requested that these officers constitute a fair cross section of the department and that past good performance records be a basis for selection. No one from the department of personnel, however, seems to have reviewed the selection to determine whether the officers were indeed "job knowledge experts" for the assigned tasks. The casual manner in which at least one of the participants was

chosen (see Burkhalter's testimony) leaves some question as to what standards the police were using. At any rate, the test maker cannot demonstrate that the consultants utilized were suited for the assignment.

While this deficiency might be considered *de minimis* had the test maker been in control of actually defining what police officers' work behaviors encompass, the latter task was left entirely to the selected officers. At trial defendants could not define the governing KSA's except to tell us what the police said they were. Counsel for defendants, Juni and Finkelman contended that it was sufficient for the police to know what the KSA's were. Neither they, nor we had to be enlightened on that issue. That proposition must surely be incorrect. It is not possible for the test maker to design an examination to measure work behaviors, or in this case KSA's, that she cannot explain, analyze or define. She must know that the KSA's identified by those with knowledge of the job actually relate to a substantial portion of the job in issue. *Procedures* would seem to require that the test designer, not the work knowledgeable consultants, be in charge of analyzing the KSA's and work behaviors in structuring the test. Here much of the test's preparation was done by various teams of police officers without the test maker understanding precisely what had been done.

The Uniform Guidelines, § 14C(2), require that a "job analysis focus on work behavior(s) and the tasks associated with them." The 71 tasks initially identified and subsequently reduced to 42 were not delineated with precision; e.g., interacts with the community and provides information and assistance to the public; performs stop and frisk of appropriate individuals; interacts with juveniles in non-arrest situations are vague definitions and do not provide the test maker with a mirror image of the police officer's work behavior. Nor was the process of identifying the KSA's sufficiently controlled to insure that they were "operationally defined" in terms of observable aspects of work behavior on the job. See Uniform Guidelines, § 14C(4) and *Procedures, passim.*

In addition, the five KSA's are not denoted with specificity; and the KSA's most imprecisely defined of all—human relations, including communication techniques—were scored as constituting 30% of a police officer's needed abilities and skills. These KSA's must entail oral communication at least in some degree. However, no oral examination was contemplated to test this aspect of the human relations KSA's. In such a case, *Procedures* suggests the KSA's must be eliminated from consideration. This was not done, and the examination must be rated unreliable by department of personnel's own yardstick.

Finally, it is not clear at what level of employment—entry or later—a police officer performs the tasks and uses the KSA's reflected in the examination. The job analysis done here did not even conform to professional standards of the department of personnel's own guide, *Procedures*. Thus the "manifest relationship" of Examination No. 8155 to the work behavior of police officers on the New York police force is, at best, highly suspect.

III

There was a consensus among the parties' experts that in operation, the examination was a paper and pencil test that required applicants to read and comprehend a set of facts and multiple choice questions about those facts and to select a correct answer from the possibilities presented. Defendants' witnesses testified that the test was designed to measure the KSA's necessary to function adequately as a police officer, and not to determine whether applicants possessed those abilities requisite to successful training for the job. Thus, it was argued that the test possessed content validity. Yet all successful applicants undergo a period of specialized training before actually assuming any duties. There was some unwillingness by defendants to concede that the tasks defined encompassed those tasks for which the successful candidate would be given training before being assigned to duty as a member of the police force, but all 42 tasks seem to speak to some aspect of the ongoing functions of policemen. Hence, the applicant must secure training in substantially all of the designated tasks.

Validation of the test as content valid is accordingly inappropriate. The Uniform Guidelines, § 14C(1) state that content validation is not appropriate when the process of selection "involves . . . abilities which an employee will be expected to learn on the job." Moreover, a content validation model may be utilized only when the test corresponds closely to those tasks that are actually performed on the job. *United States v. City of Chicago,* 573 F.2d 416, 425 (7th Cir. 1978). Neither Finkelman nor Juni contended that the test measured on the job behavior of police officers. It was agreed that, at best, the test gauged applicants' ability to comprehend and follow instructions.

Although much was made of the KSA's as identifying abilities, it is clear that the test measured constructs, not abilities or skills. While abilities are "present competence to perform an observable behavior," Uniform Guidelines, § 16A, Tr. 138, 312, constructs are intangibles, such as "intelligence, aptitude, personality, judgment." Uniform Guidelines, § 14C(1), Tr. 309. The examination attempted to test for verbal aptitude or reasoning. Both defendants' and plaintiffs' experts agreed that content validation procedures are inappropriate for determining the validity of a test for constructs. Defendants concede that the test did not measure present competence since they do not argue that the successful applicant could perform the job of police officer without first undergoing a period of training. Accordingly, the test merely identified some underlying aptitudes or propensities, and these are surely constructs. Plaintiffs' experts, Dr. Richard Barrett (Tr. 320) and Dr. James J. Kirkpatrick (Tr. 461), concluded that the test was a generalized aptitude test which measured competence to read instructions and apply them to a simulated situation.

The Uniform Guidelines, § 14C(4), require:

> to be content valid, a [test] measuring a skill or ability should either closely approximate an observable work behavior, or its product should closely approximate an observable work product . . . As the content of the [test] less resembles a work behavior, or the setting and manner of the administration of the [test] less resemble the work situation, or the result less resembles a work product, the

> less likely the [test] is to be content valid, and the greater the need for other evidence of validity.

Almost a third of the test purported to reveal the candidate human relations, including communications techniques KSA's. Yet, it is obvious that the KSA's to perform well in this area could not be tested in a written multiple choice examination. The applicant was asked to choose the best response to a simulated situation. No one could argue seriously, however, that the chosen response indicates how an applicant would actually react in real life to a situation where he must interact with real people. Judge Foley of the Northern District of New York met that contention head on in *United States v. State of New York,* 77-CV-343 (Sept. 6, 1979), and I quote his finding #175:

> The 1975 written examination for the position of trooper was a situations test which, in essence, sought "will do" responses to situations that normally do not occur behind a desk. The fact that someone selects a particular course of action as appropriate on such an examination does not mean that the same course of action would be followed by that individual under different circumstances in a real life situation. Certainly, in such situations, an individual would not consider and determine what would be the worst course of action to be followed. The situations on the written examination are simulated situations that bear little resemblance, marginal at best, to actual situations in which the behavior might occur. Sitting at a desk selecting the best alternative and the worst alternative in response to a written situation, I find, is quite different from being in a real life situation deciding what to do and doing it effectively.

See also Guardians II, *supra,* 431 F. Supp. at 433; *United States v. San Diego County, supra,* 20 E.P.D. ¶ 30,154 at 11,812.

Nor does the test fare any better in relation to the other KSA's. The questions testing the KSA's for filling out forms

were based on a situation involving the robbery of a delicatessen and arrest of the suspect. A simulated arrest form was reproduced, and the questions asked the candidate to choose relevant information to be placed in blanks on the form. Four possible choices were presented for each question. A police officer in real life, however, must select the correct form and fill in the required information without help from suggested alternatives, one of which is appropriate.

In testing the recalling details KSA's, the candidate was given ten minutes to study a booklet containing a story about a burglary and a detailed description of the suspect, vehicles, scene, etc. The booklet was collected after ten minutes, and test questions were distributed which related to a theft of $500,000 worth of gold from a jewelry company and required the recall of specific details. Being asked about details immediately after studying a text for ten minutes does not mirror the recall a police officer may have to exercise in the performance of his duties. He may have to remember details of what he saw when under the pressure of physical danger to himself or others, the risk of the escape of the suspect, while making up his mind whether to seek assistance or to use firearms, and a myriad of other crises. A police officer's memory under such critical circumstances indicates whether he has optimum recalling details KSA's in police work, and it cannot be measured in a paper and pencil test.

The questions concerning comprehending and applying sections of the law were based on 19 definitions of crimes. Each question presented a fact pattern, and the candidate was to choose from four options the crime corresponding to the fact pattern described. Again, this does not reflect the KSA's involved in a police officer's comprehending and applying sections of the law. In a real situation, the officer sees activity and must determine rather quickly whether the activity is illegal, with no definitional aids before him. He must operate on instinct and experience.

All that any of the questions tested were the KSA's involved in reading, comprehending and following instructions, not the KSA's needed to perform police work.

Defendants, in an apparent effort to meet criticism that the test was a reading test, sought to design the examination at a readability level at the 12th grade. At this level it was evidently felt that the adverse impact of the test on blacks and hispanics might be diminished or removed. Of course, where such alternatives are available and consistent with the employer's legitimate interests, they must be utilized. *See Albemarle Paper Co. v. Moody, supra,* 422 U.S. at 425.

Defendants did not altogether succeed, however, in writing an examination at the 12th grade reading level. They conceded that some of the questions, explanations or instructions required college equivalent reading skills. In any event, basing eligibility on rank order neutralized whatever gains had been achieved in lessening the adverse impact of the test on blacks and hispanics by simplifying the wording of the questions. Moreover, a test easier to read may make it possible for a larger number of blacks and hispanics to succeed, but it qualifies a larger number of whites as well, and the overall ratio of success among the three groups may remain virtually unchanged. Indeed, if one compares the distribution of scores by race on this test with that of the tests at issue in *Guardians* II, *supra,* 431 F. Supp. at 551-53 (appendix), the results are basically indistinguishable. When rank order is the basis of selection on an aptitude test such as this is, whites in the main will be clustered in the upper ranks and blacks and hispanics on lower levels. In qualifying the 12,000 highest scores, more blacks and hispanics may have been eliminated than had a pass-fail grade been set. Unless it is demonstrated that those with the highest scores will perform best as police officers, rank order as the basis of eligibility for appointment cannot be justified on the ground that this form of selection is an insurance of competence. The police department apparently presumes that the higher scorers possess more police-related KSA's, but there is nothing in the record to support that presumption.

Since the adverse impact of Examination No. 8155 has been established and defendants have failed to prove that the test has a "manifest relationship" to a police officer's duties, Title

VII has been violated. Accordingly, the examination is held to be invalid under Title VII, and defendants are permanently enjoined from using the eligibility list in rank order to make appointments to the police force.

Relief

This case is not one of first impression but involves evidence of a pattern, long continued, of discrimination against blacks and hispanics in the hiring and appointment of police officers on the New York City police force. As pointed out earlier the issue of the selection process for police officer has been in this court at least since 1972. Examinations given 1968 and 1970 were held to have an adverse impact on minority applicants. *See Guardians* II, *supra,* 431 F. Supp. at 552, 553, Tables 4 and 6. A discriminatory pattern, therefore, was found to exist in *Guardians* II and a similar pattern has been evidenced here. Moreover, in light of *Guardians* II, it would have been expected that defendants in structuring any new examinations would be especially careful to follow requisite guidelines and procedures to make certain that new tests complied with Title VII. Instead, defendants have persisted in devising and utilizing testing procedures that continue to discriminate against blacks and hispanics. Examination 8155 fails to meet even the standards established by the department of personnel itself. This studied adherence to discriminatory procedures must at this point be deemed conscious and deliberate, and insistence on maintaining procedures for the selection of police officers known to have an adverse impact on blacks and hispanics cannot be dismissed as inadvertent but must be the result of deliberate policy. Since there is no proof that police work is performed more efficiently and effectively by persons who score high on this examination and its adverse racial impact has been proven, there is no rational support for the rank order selection process.

Accordingly, I find defendants have designed an examination which they knew or should have known tests for reading skills and does not demonstrably reveal those skills that dif-

ferentiate between a good police recruit and a poor one. The test advantages whites and disadvantages blacks and hispanics without proof that it appropriately identifies those persons best suited to be police officers. *Griggs v. Duke Power Co., supra,* and cognate cases defining the scope and reach of Title VII have uniformly struck down such practices as unlawful. Defendants' assertions that the examination is a test of a policeman's skills are misleading. Moreover, testimony revealed that during the preparation phases of the test several police officers warned the test makers that the test would disfavor blacks and hispanics. I find and hold that Examination No. 8155 was designed either with a deliberate intention to discriminate against blacks and hispanics or with reckless disregard of whether the test would have that result.

While the 1973 examination for appointment to the police force has not been, as far as I know, the subject of attack, the most recent examinations prior thereto were condemned as discriminatory in *Guardians* II. In that case Dr. Bernard Cohen, testifying on the basis of a study by the Rand Institute, authorized by the parties and paid for in part by the city, to determine the significance of the distribution by race of scores on entry level examinations given in 1968 and 1970 for appointment to the New York City police force, stated that the chances of such substantial and significant differences between the distributions of the scores of whites and of blacks and hispanics occurring on a racially neutral test were "statistically less than one in a billion." *See Guardians* II, *supra,* 431 F. Supp. at 540. Now defendants have repeated the process in the current examination. The imbalance between hispanics and blacks on the one hand and whites on the other on the New York City police force is directly caused by past and current discriminatory practices. *See Association Against Discrimination in Employment v. City of Bridgeport,* 594 F.2d 306 (2d Cir. 1979).

Affirmative action is mandated as an interim measure either until such discrimination has been totally eliminated or until defendants proceed to select police officers under procedures that are in full compliance with Title VII. Requiring defen-

dants to take positive steps to eliminate the imbalance will not have adverse consequences for " 'a small number of readily identifiable' non-minority members." *Id.* at 310, citing *Kirkland v. New York State Department of Correctional Services,* 520 F.2d 420, 427, 429 (2d Cir. 1975), *rehearing en banc denied,* 531 F.2d 5 (2d Cir. 1975), *cert. denied,* 429 U.S. 823 (1976). At the January 11, 1980 hearing on relief, defendants' testimony showed that only approximately 20% of those on the eligibility list in each group called thus far had been able to meet the additional qualifications required for appointment. At this qualification rate, it is likely that all white applicants who scored 94 or higher on the test and who meet the additional requirements will be appointed, even with affirmative action being mandated.

Defendants are enjoined from using Examination No. 8155, except that defendants may use the eligibility list derived from the above examination both to achieve minority composition of the police force comparable to the labor force of the relevant hiring area, which is at least 30% black and hispanic, and to appoint during the interim 50% of their entry level officers from among qualified blacks and hispanics.

Plaintiffs are entitled to their court costs and reasonable attorneys' fees to date. Jurisdiction is retained for such further relief as may be appropriate to ensure compliance with the requirements of Title VII.

IT IS SO ORDERED.

Dated: New York, New York
January 23, 1980

ROBERT L. CARTER
U.S.D.J.

Appendix D

Order of the District Court

UNITED STATES DISTRICT COURT

SOUTHERN DISTRICT OF NEW YORK

79 Civ. 5314 RLC

THE GUARDIANS ASSOCIATION OF THE NEW YORK CITY POLICE DEPARTMENT, THE HISPANIC SOCIETY OF THE NEW YORK CITY POLICE DEPARTMENT, NYDIA I. DIAZ, JAMES MICHAEL HIDALGO, WILFRED CEBOLLERO, ANDRE LOPEZ, REINALDO SALGADO, DENISE SANTOS, DEBORAH HOLMES, and PAMELA OBEY, individually and on behalf of all those similarly situated,

Plaintiffs,

—against—

CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK, DEPARTMENT OF PERSONNEL OF THE CITY OF NEW YORK, and THE NEW YORK CITY POLICE DEPARTMENT,

Defendants.

Upon consideration of the evidence presented at the liability and relief stages of this case, and consideration of the briefs and oral arguments of the parties and *amicus curiae* United States of America and Policewomen's Endowment Association of New York City, Inc., and entry of findings of fact and conclusions of law, it is hereby ORDERED:

1. Defendants, their officers, officials, agents, employees, successors, and all persons in active concert or participation with them or any of them are hereby permanently enjoined

PRECEDING PAGE WAS BLANK

from engaging in any act or practice with respect to the selection of candidates for appointment to and training for the position in the New York City Police Department of entry-level police officer, which act or practice has the purpose or effect of discriminating against such persons because of race or national origin.

2. Defendants are hereby permanently enjoined from using Examination 8155 in any manner, except as specifically provided in paragraph 5 of this order.

3. The defendants shall seek to achieve as a long-term goal black and hispanic (hereinafter "minority") representation in the sworn ranks of the Police Department comparable to that of the minority composition of the labor force in the relevant hiring area. As of 1978, the labor force of the relevant hiring area was at least 30% black and hispanic.

4. To achieve the long-term goal set forth in paragraph 3, *supra,* the defendants shall as an interim goal appoint 50% of their entry level police officers from among qualified black and hispanic applicants. The interim hiring goal for minorities shall remain in effect until the minority representation in the sworn ranks of the Police Department is at least equal to the percentage of minorities in the labor force of the relevant hiring area as described in paragraph 3, *supra,* or until this court has found, after a hearing, that all proposed selection procedures for police officer positions have been validated in accordance with the *Uniform Guidelines on Employee Selection Procedures,* 28 C.F.R. § 50.14, 29 C.F.R. § 1607, effective September 25, 1978 ("Uniform Guidelines"), and that no further interim goals are appropriate. Nothing herein shall preclude plaintiffs from advocating the continuance of the interim goals on the basis that the continuing effects of past discrimination have not been eliminated.

5. To satisfy the goals set forth above in paragraphs 3 and 4, the defendants may use an eligibility list derived from Exami-

nation No. 8155 as the pool from which it selects police officers. At such time as that eligibility list or any future eligibility list for the position of police officer does not contain sufficient minority candidates to meet the interim goals set out in paragraph 4, *supra,* the City shall take whatever steps are necessary to achieve the interim hiring goals.

6. Defendants shall make reasonable efforts to develop within a reasonable period of time a procedure for the selection of candidates for the entry level position of police officer which shall be lawful and validated in accordance with the Uniform Guidelines or successor guidelines similarly promulgated, and which is consistent with generally accepted psychological standards as defined by the American Psychological Association from time to time, and which has the least adverse impact on minority applicants. Consistent with this requirement defendants (a) shall examine all reasonably available alternative selection procedures on the subject of testing of police officer applicants, and (b) shall consult with industrial psychologists, psychometricians and/or others who have experience in the field of selection testing, and preferably who have performed or have knowledge of analyses of the job of police officers.

7. Defendants may continue to use the current qualifications and selection criteria for police officer positions. However, no such qualification or selection criterion shall be a valid basis of or defense for failure to meet the interim hiring goals set out in paragraph 4, *supra,* unless the court has ruled that such qualification or selection criterion has been validated in accordance with the Uniform Guidelines and has determined that there is no further basis for continuing the interim goals.

8. Members of the plaintiff class shall be afforded the opportunity, at appropriate later proceedings, to show that they are entitled to an award of back pay and constructive seniority.

9. Plaintiffs are entitled to their court costs and reasonable attorneys' fees to date. The amount of such costs and fees shall be set by the court after a hearing. Costs and attorneys' fees for work done in the future shall be fixed in such manner as the court may determine.

10. Within thirty (30) days after the entry of this order, and every six (6) months thereafter, the defendant city shall submit to the plaintiffs the following reports:

(a) A list of its then current uniformed employees in the Police Department showing for each person: name, address, race or national origin, police station or other place of assignment, date of appointment, rank, and date such rank was achieved.

(b) The total number of uniformed personnel employed by the Police Department, by rank, race and national origin.

(c) A list of all minority applicants for all vacancies, including date of application, names, addresses, and telephone numbers, whether the applicant was accepted or rejected and reason(s) for rejection.

(d) The name, address, and telephone number of any minority employee involuntarily terminated prior to the completion of the probationary period, and reason(s) for termination.

(e) List of all hires, promotions and voluntary and involuntary terminations showing race and national origin.

11. Defendants shall provide to plaintiffs, upon request, such other records or documents as are necessary to monitor compliance with this order.

12. The court retains jurisdiction of this action for such further relief or other orders as may be necessary or appropri-

ate to enforce and insure rights to equal employment opportunity within the New York City Police Department.

SO ORDERED.

Dated: New York, New York
January 11, 1980

ROBERT L. CARTER
U.S.D.J.